DOCUMENT RESUME

ED 253 593 TM 850 143

AUTHOR	Schrader, William B.
TITLE	Three Studies of SAT-Verbal Item Types.
INSTITUTION	College Entrance Examination Board, New York, N.Y.; Educational Testing Service, Princeton, N.J.
REPORT NO	CB-R-84-7; ETS-RR-84-33
PUB DATE	84
NOTE	59p.; Small print in some tables.
AVAILABLE FROM	College Board Publications, Box 886, New York, NY 10101 ($[illegible].00).
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Achievement Tests; Class Rank; *College Entrance Examinations; Correlation; Difficulty Level; Grade Point Average; Higher Education; High Schools; Reading Comprehension; *Test Items; *Test Validity; *Verbal Tests
IDENTIFIERS	Analogies; Antonyms; *Scholastic Aptitude Test; Sentence Completion Method

ABSTRACT

Each of the three studies in this report examines a different aspect of the basic question of what the four item types (analogies, antonyms, reading comprehension, and sentence completion) on the Scholastic Aptitude Test (SAT) verbal section are measuring and of whether a change in the relative emphasis on the various item types would enhance the validity of SAT-verbal scores. The first study, "A Survey of Item and Test Analysis Results for SAT-Verbal Item Types," examines differences in difficulty and biserial correlations between the various types, and differences in patterns of attempts from one item type to another that may have resulted from differences in candidate strategies. The second study, "The Relation of SAT Reading and Vocabulary Scores to Measures of High School Performance," provides information on the correlations of SAT reading and vocabulary scores with College Board Achievement Test scores, with self-reported high school grades, and with high school rank, and on multiple correlations when each subscore is combined with SAT-mathematical scores and high school rank. The third study, "The Validity of SAT-Verbal Item Types," includes validity coefficients and multiple correlations for the four item types when scores are based on all items of each type and also when scores are based on matched subsets including eight items of each type. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

S. W. Gardner

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

College Board Report

No. 84-7

Three Studies of SAT-Verbal Item Types

William B. Schrader

Three Studies of SAT-Verbal Item Types

William B. Schrader

College Board Report No. 84-7

ETS RR No. 84-33

College Entrance Examination Board, New York, 1984

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101. The price is $5.

Copyright © 1984 by College Entrance Examination Board. All rights reserved.
The "College Board" and the acorn logo are registered trademarks of the College Entrance Examination Board.
Printed in the United States of America.

CONTENTS

Preface 1

A Survey of Item and Test Analysis Results for SAT-Verbal Item Types 3

Abstract 3

Introduction 3

Results: Item Difficulty and Examinee Nonresponse 3

Results: Equated Deltas and Biserial Correlations 14

Results: True-Score Correlations 14

Summary and Conclusions 22

Tables

1. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for Three Recent Forms 4
2. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for High and Low Fifths on Three Recent Forms 4
3. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for Forms 3CSA5, 3CSA6, and 3CSA7 Separately 5
4. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for High and Low Fifths for Forms 3CSA5, 3CSA6, and 3CSA7 Separately 6
5. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section I: Results for Three Recent Forms 7
6. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section I: Results for High and Low Fifths on Three Recent Forms 8
7. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section II: Results for Three Recent Forms 9
8. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section II: Results for High and Low Fifths on Three Recent Forms 10
9. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for Four Black Groups Combined and Four White Groups Combined 11
10. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for Four Black Groups Separately and Four White Groups Separately 11
11. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section I: Results for Four Black Groups Combined and Four White Groups Combined ... 12
12. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of the SAT-Verbal Section II: Results for Four Black Groups Combined and Four White Groups Combined .. 13
13. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Frequency of Each Item Type for Three Recent Forms Combined 15

14. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Percentage of Each Item Type for Three Recent Forms Combined 16

15. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Percentage of Each Item Type for Form 3CSA5 17

16. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for SAT-Verbal Item Types: Percentage of Each Item Type for Form 3CSA6 18

17. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Percentage of Each Item Type for Form 3CSA7 19

18. Frequency Distribution of Equated Deltas by Item Type of Three Recent Forms of SAT-Verbal Sections 20

19. Frequency Distribution of Biserial Correlation by Item Type of Three Recent Forms of SAT-Verbal Sections 20

20. True-Score Correlations of SAT-Verbal Item Types for 20 SAT Forms 21

The Relation of SAT Reading and Vocabulary Scores to Measures of High School Performance 23

Abstract 23

Introduction 23

Method 24

Results 24

Summary and Conclusions 30

References 30

Tables

1. Intercorrelations of SAT Scores and High School Rank (Based on 849,175 Examinees in 1981 Senior File) 24

2. Multiple Correlation with High School Rank of Combinations of SAT Scores (Based on 849,175 Examinees in 1981 Senior File) 24

3. Correlation of SAT Scores and High School Rank with Achievement Test Scores 25

4. Multiple Correlation of Various Combinations of SAT Scores and High School Rank with Achievement Test Scores 26

5. Correlation of Test Scores with Course Grades for Students Taking Each Achievement Test 26

6. Multiple Correlation with Course Grade of Combination of SAT Scores for Students Taking Each Achievement Test 27

7. Multiple Correlation with Course Grade of Various Combinations of SAT Scores and Achievement Test Scores for Students Taking Each Achievement Test 28

8. Mean Score on Reading and Vocabulary for Students Taking Each Test 28

9. Correlation of Reading, Vocabulary, and SAT-V with Measures of Ability in English and with High School Rank 29

10. Correlation with English Composition Test Essay and Objective Scores with SAT Scores (Based on 55,673 Students Who Took the English Composition Test in December 1977) 29

The Validity of SAT-Verbal Item Types 31

Abstract 31

Purpose 31

Background 31

Data Base 32

Constructing the Matching Subsets 32

Analysis Methods 33

Results 33

Discussion and Conclusions 44

References 45

Tables

1. Raw Deltas of Items Having 85 Percent or More Attempts and 8-Item Subsets of Each Type Matched on Difficulty Level 32

2. Correlations with First-Year Average Grades of Scores on Matched Subsets of Items and on All Items of Each Item Type 34

3. Correlations of Scores Based on Matched Subsets of Items with First-Year Average Grades in Each College 35

4. Correlations of Scores Based on All Items with First-Year Average Grades in Each College 36

5. Intercorrelations of Scores Based on Matched Item-Type Subsets, Scores Based on All Items, SAT-Verbal, SAT-Mathematical, High School Record, and First-Year Average Grades (Based on Pooled Standardized Covariances for 48 Colleges) 38

6. Intercorrelations of Scores Based on Matched Item-Type Subsets, Scores Based on All Items, SAT-Verbal, SAT-Mathematical, High School Record, and First-Year Average Grades, for Men and Women Separately (Mean Equally-Weighted Standardized Covariances) 39

7. Intercorrelations of Scores Based on Matched Item-Type Subsets, Scores Based on All Items, SAT-Verbal, SAT-Mathematical, High School Record, and First-Year Average Grades for Seven Colleges with Relatively High SAT-V Means and Seven Colleges with Relatively Low SAT-V Means (Based on Pooled Standardized Covariances) 40

8. Correlations of SAT Scores and High School Record with First-Year Average Grades in Each College 41

9. Multiple Correlations with First-Year Average Grades in Each College Using Item-Type Scores Based on Matched Subsets of Items 42

10. Multiple Correlations with First-Year Average Grades in Each College Using Item-Type Scores Based on All Items 43

Preface

Since 1952, the SAT-verbal sections have been composed of the same four item types: analogies, antonyms, reading comprehension, and sentence completion. Beginning in the 1974–75 testing years, two verbal subscores have been reported: vocabulary, based on analogies and antonyms items, and reading, based on reading comprehension and sentence completion items. In 1981, a summary of validity coefficients for vocabulary and reading subscores prepared by Ramist indicated that the vocabulary items were making a surprisingly small contribution to the correlation of SAT-verbal scores with college grades. The three studies included in this report were designed to obtain a better understanding of Ramist's results and to examine the possible implications of statistical results on the four item types for planning future forms of the SAT-verbal sections. Each of the three studies examined a different aspect of the basic questions of what the four SAT-verbal item types and the two SAT-verbal subscores are measuring and of whether a change in the relative emphasis on the various item types would enhance the validity of SAT-verbal scores.

The first study, "A Survey of Item and Test Analysis Results for SAT-Verbal Item Types," throws light on differences in difficulty and biserials between the various item types, on how the item types are used in current test forms in meeting test specifications, and on differences in patterns of attempts from one item type to another that may have resulted from differences in candidate strategies. Evidence on these topics needs to be considered in evaluating possible changes in the test.

The second study, "The Relation of SAT Reading and Vocabulary Scores to Measures of High School Performance," provides information on the correlations of reading and vocabulary with scores on the College Board Achievement Tests, with self-reported high school grades in various academic fields, and with high school rank, and on multiple correlations when each subscore is combined with SAT-mathematical scores and high school rank. Results indicate that vocabulary and reading are measuring similar abilities, but that reading has a higher correlation than does vocabulary with most measures of high school performance.

The third study, "The Validity of SAT-Verbal Item Types," includes validity coefficients and multiple correlations for the four items types when scores are based on all items of each type and also when scores are based on matched subsets including eight items of each type. The matched subsets of items were approximately equal in level and spread of difficulty for all item types and included only items that were answered by 85 percent or more of the examinees in the sample. The results of this study must be regarded as tentative, particularly because only one form of the SAT-verbal sections could be studied. However, although reading comprehension has somewhat lower validity than the other three item types when matched subsets are compared, the results tend to support the view that the four item types are about equal in validity.

A Survey of Item and Test Analysis Results for SAT-Verbal Item Types

ABSTRACT

A recent finding that the reading subscores on the SAT-verbal sections have substantially higher validity than the vocabulary subscores has stimulated interest in the four item types of which the SAT-verbal sections are composed. This report provides a summary of data from item and test analyses on the difficulty level and on patterns of nonresponse for the four item types, on the extent to which each item type supplies items having both a relatively high difficulty level and a reasonably high biserial correlation, and on the true-score intercorrelations of the four item types.

The comparisons of the four item types included in this survey provide a description of how these item types are used in meeting test specifications for the SAT-verbal sections. Generalizations about the characteristics of the item types should therefore be considered tentative. The fact that each item type was not administered as a separately timed subtest also tends to limit the generality of the findings. In the test forms studied, reading comprehension provided a smaller proportion of items that have both a relatively high difficulty level and a relatively high biserial correlation than did the other three item types. Antonyms and analogies provided a larger proportion of highly difficult items but were also more likely to include items having relatively low biserial correlations than the other two item types.

The study also included a summary of true-score correlations. Results of these analyses indicate that sentence completion subscores are about as closely related to analogies and antonyms as they are to reading comprehension subscores in the 20 recent test forms included in the survey.

INTRODUCTION

A recent finding that the reading subscores of SAT-verbal sections have substantially higher validity coefficients with college grades than the vocabulary subscores[1] has stimulated interest in possible changes in the content and format of the SAT-verbal sections. In particular, the fact that the reading subscore is based on reading comprehension and sentence completion items and the vocabulary subscore is based on antonyms and analogies items suggests that there is a need for studying the statistical characteristics of each of the four item types. Moreover, the fact that the reading and vocabulary subscores are not based on separately timed sections suggests that the relation of difficulty level and of examinee nonresponse to item type and to the position of an item in the test deserves careful study. Still another important question concerns the extent to which each item type is supplying difficult items that have a reasonably high biserial correlation. Items having these characteristics are needed in order to fulfill test specifications.

The purpose of this survey is to summarize data from item and test analyses that are relevant to these topics. The results of this survey deserve consideration not only because they are relevant to decisions about a possible restructuring of the SAT-verbal sections but also because they throw light on possible problems in interpreting validity findings for subscores that are not based on separately timed sections. Data from completed item and test analyses of the SAT-verbal sections provide an excellent source of data about important statistical characteristics of these item types. By summarizing data from these sources, it is possible to compare the four item types with respect to their difficulty level and the strategies that examinees adopt in responding or not responding to the items. Separate comparisons will also be made based on high-scoring and low-scoring examinees and on black and white examinees. In addition, data will be summarized on the relation between item difficulty and the biserial correlation between item performance and total score and on similarities and differences of items types in the abilities that they measure.

Evidence of the kind obtained in this study is unlikely to provide a definitive basis for decisions about the content and format of the SAT-verbal sections. Its value depends on its contribution when it is taken into account in considering possible implications of validity studies.

RESULTS: ITEM DIFFICULTY AND EXAMINEE NONRESPONSE

Standard item analyses for three recent forms provide the data for the first set of results. The forms included were first administered in October 1980 (3CSA5), November 1980 (3CSA6), and December 1980 (3CSA7). Sample sizes for these analyses vary somewhat (from 1650 to 1790). In the summary figures, equal weight is given to each sample. Characteristics to be summarized include: percentage of

1. L. Ramist, "Validity of SAT-Verbal Subscores," Memorandum, February 12, 1981.

items answered correctly after correction for chance success, percentage of items omitted, percentage of items not reached, and percentage of items not attempted. (Omitted items *precede* the last item in a separately timed section; items not reached *follow* the last item attempted in the section.)

It was considered desirable to adjust the percentage correct for chance success so that the measure of item difficulty level could be assumed to be relatively little affected by differences in nonresponse. For this measure of item difficulty, it can be assumed that the extent to which examinees guess will have relatively little effect on the results. Accordingly, if nonresponse is greater for more difficult items than for easier items, the relationship can presumably be attributed to a tendency for examinees not to respond to more difficult items rather than resulting in part from the fact that nonresponse tends to lower the percentage of right answers.

Table 1. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for Three Recent Forms. Note: Percent Correct Adjusted for Chance Success

Item Number	*Item Type*	*Correct*	*Omitted*	*Not Reached*	*Not Attempted*
1–15	Antonyms	42.4	15.5	0.0	15.5
16–20	Sentence Completion	49.0	4.5	0.0	4.5
21–30	Reading Comprehension	43.8	6.6	0.2	6.9
31–35	Sentence Completion	36.7	7.7	2.7	10.3
36–45	Analogies	37.7	6.4	17.9	24.2
1–10	Antonyms	41.2	14.2	0.0	14.2
11–15	Sentence Completion	52.7	4.2	0.0	4.2
16–25	Analogies	46.9	12.9	0.0	12.9
26–40	Reading Comprehension	32.1	4.7	7.8	12.5

In the forms of the SAT-verbal sections included in this summary, nine sets of items, each set composed of items of the same type, are presented. (In this report, a sequence of items of the same type will be called a segment.) Results for the nine segments of the SAT-verbal sections shown in Table 1 indicate that percentage correct tends to be lowest for the last two segments of the first section and for the last segment of the second section. Because percentage correct has been adjusted to take account of chance success by subtracting one-fourth of the percentage of wrong answers from the percentage of right answers, it is unlikely that the fairly large number of unanswered items in those segments contributes to this result. It is plausible, however, that examinees would work less carefully as they approach the end of each separately timed section because of time pressure. The evaluation of this hypothesis would require further research relating item position to item difficulty and biserial correlation. Leary and Dorans (1982) have reviewed the literature on the relation of item rearrangement to test performance. The results of the present study suggest that the position of an item within its segment as well as its position within the test may affect its statistical characteristics.

Results for omitted items show that antonyms and analogies tend to have noticeably higher percentages of omits than reading comprehension and sentence completion items. Results for omits on the analogies segment at the end of the first section do not show a higher percentage of omits, probably because the percentage of items not reached is high for this segment.

Table 2 shows results for each segment for examinees in the high fifth and the low fifth on the SAT-verbal sections. The tendency for percentage correct to fall off toward the

Table 2. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for High and Low Fifths on Three Recent Forms

		Correct		*Omitted*		*Not Reached*		*Not Attempted*	
Item Number	*Item Type*	*High Fifth*	*Low Fifth*	*High Fifth*	*Low Fifth*	*High Fifth*	*Low Fifth*	*High Fifth*	*Low Fifth*
1–15	Antonyms	67.3	18.7	8.9	18.3	0.0	0.0	8.9	18.3
16–20	Sentence Completion	75.7	25.4	1.6	6.3	0.0	0.0	1.6	6.3
21–30	Reading Comprehension	70.4	17.5	2.2	10.3	0.0	1.0	2.2	11.3
31–35	Sentence Completion	69.2	10.1	2.5	10.9	0.1	7.9	2.6	18.8
36–45	Analogies	61.2	14.2	5.4	6.5	8.0	29.0	13.4	35.4
1–10	Antonyms	66.3	15.0	7.6	18.3	0.0	0.0	7.6	18.3
11–15	Sentence Completion	80.8	23.7	1.8	6.0	0.0	0.0	1.8	6.0
16–25	Analogies	71.2	21.9	7.0	15.2	0.0	0.0	7.0	15.2
26–40	Reading Comprehension	62.6	8.0	2.0	6.9	2.2	12.9	4.2	19.8

Table 3. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for Forms 3CSA5, 3CSA6, and 3CSA7 Separately

Item Number	*Item Type*	*Group*	*Correct*	*Omitted*	*Not Reached*	*Not Attempted*
1–15	Antonyms	3CSA5	45.3	16.5	0.0	16.5
		3CSA6	45.4	14.5	0.0	14.5
		3CSA7	36.7	15.6	0.0	15.6
16–20	Sentence Completion	3CSA5	50.5	4.6	0.0	4.6
		3CSA6	40.3	4.4	0.0	4.4
		3CSA7	56.3	4.6	0.0	4.6
21–30	Reading Comprehension	3CSA5	48.8	4.1	0.3	4.3
		3CSA6	48.4	5.1	0.3	5.4
		3CSA7	34.3	10.7	0.2	10.9
31–35	Sentence Completion	3CSA5	26.7	7.7	2.6	10.4
		3CSA6	39.8	7.2	4.0	11.3
		3CSA7	43.5	8.0	1.4	9.4
36–45	Analogies	3CSA5	37.7	4.4	17.4	21.8
		3CSA6	40.2	8.7	19.1	27.8
		3CSA7	35.0	6.0	17.1	23.1
1–10	Antonyms	3CSA5	41.4	19.6	0.0	19.6
		3CSA6	45.3	11.4	0.0	11.4
		3CSA7	37.0	11.6	0.0	11.6
11–15	Sentence Completion	3CSA5	45.6	4.5	0.0	4.5
		3CSA6	70.1	2.3	0.0	2.3
		3CSA7	42.4	5.8	0.0	5.8
16–25	Analogies	3CSA5	51.5	12.7	0.0	12.7
		3CSA6	48.8	13.6	0.0	13.6
		3CSA7	40.2	12.4	0.0	12.4
26–40	Reading Comprehension	3CSA5	32.3	5.5	6.0	11.5
		3CSA6	31.4	5.7	5.2	10.9
		3CSA7	32.6	2.8	12.2	15.0

Table 4. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for High and Low Fifths for Forms 3CSA5, 3CSA6, and 3CSA7 Separately

Item Number	Item Type	Group	Correct		Omitted		Not Reached		Not Attempted	
			High Fifth	Low Fifth	High Fifth	Low Fifth	High Fifth	Low Fifth	High Fifth	Low Fifth
1–15	Antonyms	3CSA5	69.8	24.3	8.4	19.3	0.0	0.0	8.4	19.3
		3CSA6	71.6	19.0	8.8	17.1	0.0	0.0	8.8	17.1
		3CSA7	60.6	13.0	9.6	18.6	0.0	0.0	9.6	18.6
16–20	Sentence Completion	3CSA5	77.3	24.5	1.5	5.9	0.0	0.0	1.5	5.9
		3CSA6	71.7	16.0	1.5	6.3	0.0	0.0	1.5	6.3
		3CSA7	78.0	35.7	1.8	6.7	0.0	0.0	1.8	6.7
21–30	Reading Comprehension	3CSA5	73.4	18.9	1.3	6.7	0.0	1.0	1.3	7.7
		3CSA6	74.2	22.3	1.8	7.8	0.0	1.1	1.8	8.8
		3CSA7	63.6	11.2	3.5	16.5	0.0	0.9	3.5	17.4
31–35	Sentence Completion	3CSA5	62.3	0.7	1.6	10.2	0.0	8.4	1.6	18.6
		3CSA6	76.5	11.5	2.4	10.2	0.2	11.2	2.6	21.4
		3CSA7	68.8	18.2	3.6	12.1	0.0	4.1	3.6	16.2
36–45	Analogies	3CSA5	62.7	15.7	3.4	4.9	5.5	31.8	8.9	36.7
		3CSA6	63.1	16.0	7.0	8.3	7.1	34.0	14.2	42.3
		3CSA7	57.9	10.9	5.9	6.3	11.3	21.0	17.2	27.3
1–10	Antonyms	3CSA5	63.8	18.9	11.8	23.5	0.0	0.0	11.8	23.5
		3CSA6	71.3	15.5	3.7	18.2	0.0	0.0	3.7	18.2
		3CSA7	63.9	10.6	7.2	13.4	0.0	0.0	7.2	13.4
11–15	Sentence Completion	3CSA5	81.3	13.7	1.4	6.4	0.0	0.0	1.4	6.4
		3CSA6	90.0	42.7	0.8	4.8	0.0	0.0	0.8	4.8
		3CSA7	71.1	14.8	3.2	6.9	0.0	0.0	3.2	6.9
16–25	Analogies	3CSA5	73.4	27.5	5.9	15.6	0.0	0.0	5.9	15.6
		3CSA6	74.1	25.5	6.9	16.9	0.0	0.0	6.9	16.9
		3CSA7	66.1	12.5	8.3	13.1	0.0	0.1	8.3	13.2
26–40	Reading Comprehension	3CSA5	64.9	8.4	2.3	8.0	1.8	9.3	4.1	17.3
		3CSA6	61.7	5.9	2.3	9.0	1.0	9.5	3.3	18.5
		3CSA7	61.2	9.6	1.4	3.7	3.6	19.8	5.1	23.5

Table 5. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section I: Results for Three Recent Forms

Item Number	*Item Type*	*Correct*	*Omitted*	*Not Reached*	*Not Attempted*
1	Antonyms	88.4	1.2	0.0	1.2
2	"	82.6	1.3	0.0	1.3
3	"	84.5	1.5	0.0	1.5
4	"	77.8	4.7	0.0	4.7
5	"	54.4	11.2	0.0	11.2
6	"	58.2	8.0	0.0	8.0
7	"	54.2	12.4	0.0	12.4
8	"	48.3	11.7	0.0	11.7
9	"	28.5	23.5	0.0	23.5
10	"	18.3	22.0	0.0	22.0
11	"	19.0	21.9	0.0	21.9
12	"	8.4	27.9	0.0	27.9
13	"	4.4	28.4	0.0	28.4
14	"	5.1	35.6	0.0	35.6
15	"	4.5	21.7	0.0	21.7
16	Sentence Completion	83.4	0.5	0.0	0.5
17	"	77.6	0.9	0.0	0.9
18	"	47.8	4.0	0.0	4.0
19	"	31.2	5.7	0.0	5.7
20	"	5.2	11.4	0.0	11.4
21	Reading Comprehension	69.2	2.0	0.0	2.0
22	"	55.8	2.4	0.0	2.5
23	"	52.5	6.5	0.0	6.5
24	"	53.2	3.3	0.0	3.3
25	"	45.6	3.3	0.1	3.4
26	"	52.1	5.6	0.2	5.8
27	"	28.1	7.9	0.3	8.1
28	"	43.5	8.2	0.4	8.6
29	"	12.4	14.1	0.6	14.7
30	"	26.0	12.8	0.9	13.7
31	Sentence Completion	71.8	2.6	1.4	3.9
32	"	58.6	4.7	1.9	6.6
33	"	33.7	7.5	2.6	10.0
34	"	12.8	14.4	3.5	17.8
35	"	6.5	9.2	4.1	13.3
36	Analogies	84.2	0.5	5.1	5.6
37	"	69.2	2.3	6.1	8.4
38	"	71.8	3.6	7.0	10.6
39	"	54.1	2.8	8.2	11.0
40	"	55.8	3.7	10.1	13.8
41	"	18.8	13.1	16.5	29.6
42	"	8.9	17.2	20.5	37.8
43	"	3.4	13.0	24.2	37.3
44	"	8.6	7.5	32.5	40.0
45	"	1.9	0.0	48.2	48.2

end of the test seems somewhat more pronounced for the low fifth than for the high fifth, particularly for the sentence completion segment in the first section. With respect to omitted items, antonyms and analogies show a noticeably higher percentage than the other two item types, except for the analogies segment at the end of the first section, a result which is consistent with the findings for the total group.

Table 3 shows overall results and Table 4 shows results for the high and low fifth for each of the three forms separately. Although there are some divergent results, particularly the high percentage correct for the sentence completion items in the second section of form 3CSA6, the patterns for the three forms seem reasonably similar.

Tables 5 through 8 present results for items in each position in the test for the three forms combined. A rather striking pattern seems apparent in these tables. Items which are placed near the end of a segment tend to show markedly lower percentages of correct answers and markedly higher percentages of omits in the first part of the section and items not attempted in the latter part of the section. Presumably, this pattern is mainly the result of the fact that items are arranged in ascending order of difficulty within each segment. However, the results also suggest that examinees may skip to the next segment when they reach the more difficult items or they may hurry through the last portion of the segment in order to spend their time on the easier items at the beginning of the next segment. To the extent that this effect occurs, segments in any part of the test may in fact be speeded for some examinees, resulting in possible overestimates of reliability. Moreover, the relative allocation of effort among the four item types may be affected by the perceived difficulty of the items, and this, in turn, might affect the relative validity of the item types.

Tables 9 through 12 are based on four item analyses. In each of the four, the sample included 1,000 students. For two of the forms (WSA4 and WSA5) white and black samples were drawn in the usual ways from the total examinee group. For the third form (XSA2) separate male and female samples were selected so that black and white groups were approximately matched in verbal ability on the basis of their scores on a 40-item verbal equating section. As a result, black–white differences in performance on the operational test are substantially reduced for these samples. It was judged that results for these samples could be pooled with those for the other two forms for comparisons of the four item types with each other, although the pooled sample cannot be regarded as a typical cross section of examinees. In pooling data from the four analyses, each of the two XSA2 samples was given half the weight of the other two samples, so that each of the three test forms would receive equal weight.

Table 9 shows results for each of the eight segments of which the SAT-verbal sections were composed when these forms were administered. Results for percentage correct indicate that, on the average for the three forms, the reading

Table 6. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section I: Results for High and Low Fifths on Three Recent Forms

Item Number	Item Type	Correct		Omitted		Not Reached		Not Attempted	
		High Fifth	Low Fifth	High Fifth	Low Fifth	High Fifth	Low Fifth	High Fifth	Low Fifth
1	Antonyms	98.6	64.5	0.1	2.9	0.2	0.0	0.3	2.9
2	"	97.6	53.2	0.1	4.3	0.0	0.0	0.1	4.3
3	"	98.0	54.8	0.0	4.4	0.0	0.0	0.0	4.4
4	"	95.4	44.4	0.8	11.2	0.0	0.0	0.8	11.2
5	"	89.5	16.0	2.6	18.2	0.0	0.0	2.6	18.2
6	"	87.8	22.5	1.1	16.2	0.0	0.0	1.1	16.2
7	"	89.2	18.3	3.7	20.3	0.0	0.0	3.7	20.3
8	"	87.8	12.6	1.9	15.4	0.0	0.0	1.9	15.4
9	"	56.9	9.0	12.7	25.6	0.0	0.0	12.7	25.6
10	"	52.9	−1.6	10.4	25.3	0.0	0.0	10.4	25.3
11	"	52.0	0.8	10.0	27.1	0.0	0.0	10.0	27.1
12	"	28.8	−0.8	24.1	25.0	0.0	0.0	24.1	25.0
13	"	29.2	−5.5	20.2	27.3	0.0	0.0	20.2	27.3
14	"	24.2	−1.4	31.1	32.5	0.0	0.0	31.1	32.5
15	"	22.2	−5.6	15.3	19.6	0.0	0.0	15.3	19.6
16	Sentence Completion	97.3	59.5	0.1	1.7	0.0	0.0	0.1	1.7
17	"	98.0	50.1	0.1	1.3	0.0	0.0	0.1	1.3
18	"	79.2	24.0	1.6	5.2	0.0	0.0	1.6	5.2
19	"	68.0	2.4	2.0	8.9	0.0	0.0	2.0	8.9
20	"	35.8	−9.1	4.2	14.5	0.0	0.0	4.2	14.5
21	Reading Comprehension	87.4	42.7	0.6	4.2	0.0	0.0	0.6	4.2
22	"	80.5	19.7	0.8	5.8	0.0	0.1	0.8	5.9
23	"	76.1	29.1	1.8	11.6	0.0	0.2	1.8	11.8
24	"	78.9	21.7	0.8	6.2	0.0	0.2	0.8	6.4
25	"	70.9	18.8	1.0	5.7	0.0	0.3	1.0	6.0
26	"	76.9	24.9	1.3	9.7	0.0	0.8	1.3	10.5
27	"	59.2	3.8	2.8	11.1	0.0	1.0	2.8	12.0
28	"	78.1	14.3	2.5	12.2	0.0	1.5	2.5	13.6
29	"	45.3	−5.0	5.6	18.9	0.0	2.4	5.6	21.3
30	"	50.6	4.7	5.1	18.0	0.0	3.5	5.1	21.5
31	Sentence Completion	95.9	36.2	0.1	5.2	0.0	4.9	0.1	10.1
32	"	87.3	26.1	1.0	7.7	0.0	6.7	1.0	14.4
33	"	75.6	−0.3	1.6	13.2	0.1	7.7	1.7	20.9
34	"	47.3	−3.2	6.9	14.9	0.1	9.2	7.0	24.1
35	"	40.1	−8.0	3.0	13.3	0.1	11.0	3.1	24.2
36	Analogies	97.1	54.6	0.0	1.5	0.5	13.2	0.5	14.7
37	"	92.4	33.2	0.1	6.0	0.7	15.5	0.8	21.5
38	"	95.2	33.1	0.2	7.6	0.9	17.9	1.1	25.6
39	"	86.3	19.0	0.9	4.0	1.2	20.2	2.1	24.2
40	"	89.6	14.7	1.1	6.7	1.4	23.6	2.4	30.3
41	"	50.9	−3.1	8.0	13.3	3.7	29.9	11.6	43.2
42	"	36.7	−3.6	16.8	12.5	6.9	34.0	23.6	46.5
43	"	20.0	−2.5	15.4	8.3	10.6	37.6	26.1	46.0
44	"	28.2	−0.8	11.7	4.9	18.4	43.7	30.1	48.6
45	"	15.4	−2.4	0.0	0.0	35.8	53.8	35.8	53.8

Table 7. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section II: Results for Three Recent Forms

Item Number	*Item Type*	*Correct*	*Omitted*	*Not Reached*	*Not Attempted*
1	Antonyms	80.0	1.5	0.0	1.5
2	"	73.2	4.0	0.0	4.0
3	"	69.4	8.4	0.0	8.4
4	"	65.5	6.8	0.0	6.8
5	"	52.7	9.2	0.0	9.2
6	"	40.2	17.0	0.0	17.0
7	"	11.3	23.9	0.0	23.9
8	"	19.5	19.7	0.0	19.7
9	"	0.9	28.2	0.0	28.2
10	"	−0.4	23.1	0.0	23.1
11	Sentence Completion	80.3	0.5	0.0	0.5
12	"	72.8	1.6	0.0	1.6
13	"	60.2	3.6	0.0	3.6
14	"	38.2	5.8	0.0	5.8
15	"	12.1	9.5	0.0	9.5
16	Analogies	83.3	0.9	0.0	0.9
17	"	76.5	2.3	0.0	2.3
18	"	76.9	1.1	0.0	1.1
19	"	68.9	3.4	0.0	3.4
20	"	48.3	10.2	0.0	10.2
21	"	57.7	5.7	0.0	5.7
22	"	29.8	22.9	0.0	22.9
23	"	12.7	19.9	0.0	19.9
24	"	12.5	37.8	0.0	37.8
25	"	1.9	24.8	0.1	24.8
26	Reading Comprehension	35.2	3.9	0.1	4.0
27	"	39.4	4.5	0.2	4.7
28	"	19.1	7.1	0.2	7.3
29	"	59.7	5.1	0.6	5.7
30	"	42.1	3.9	0.8	4.7
31	"	43.9	4.3	1.1	5.5
32	"	34.1	3.3	2.4	5.7
33	"	34.4	4.0	3.0	7.0
34	"	21.1	5.2	4.7	9.9
35	"	33.3	8.6	6.1	14.7
36	"	45.5	6.7	10.3	17.0
37	"	7.4	7.9	16.0	24.0
38	"	29.4	3.8	18.7	22.5
39	"	25.8	2.0	22.5	24.5
40	"	11.3	0.0	30.1	30.1

comprehension segment in the first section was relatively easy for both black and white examinees and that the analogies segment in the second section was the most difficult of the eight for both black and white examinees. The antonyms and analogies segments of the first section and the antonyms segment of the second section had noticeably more items omitted than the other segments. In the second section, the analogies segment appeared last. It showed a relatively low percentage of items omitted but a high percentage of items not attempted. Results for percentages omitted are similar for black and white examinees, but black examinees show a noticeably higher percentage of items not reached than white examinees.

Table 10 shows results for each of the four forms separately. Because the two samples for form XSA2 were drawn so as to reduce black–white differences in score level, only results within each sample may be considered directly comparable. With respect to percentage correct, the reading comprehension segment at the end of the first section is relatively easy for WSA4 and WSA5, but not for XSA2. The analogies segment at the end of the second section shows either the lowest or next to the lowest percentage correct among the eight segments for every sample. Comparisons of items omitted and items not reached across the eight segments reveal generally similar patterns across the four samples although, as would be expected, the particular items included in each form appear to affect the results. For example, sentence completion in form WSA4 shows an unusually large number of omitted items. As in Table 9, results for omits are reasonably similar for black and white examinees, and black examinees show a high percentage of items not reached.

Tables 11 and 12 show results for each item position based on average results for the four samples. As in the results shown in Tables 5 through 8, there is clear evidence that percentage correct tends to fall off toward the end of each segment. Whether this trend is entirely attributable to an increase in item difficulty or results in part from a tendency for some examinees to skip or hurry through the more difficult items at the end of each segment cannot be determined from the available data. Results for items omitted on the first three segments of each section (for which percentage not reached is low) suggest that the tendency to omit the more difficult items at the end of a segment is more pronounced for antonyms and analogies than for reading comprehension or sentence completion. Examinees may, of course, find it easier to judge relative difficulty for antonyms and analogies items than for the other two item types.

In general, the results of the item analysis summaries indicate that the organization of the test into segments and the practice of arranging items within each segment in generally ascending order of difficulty produce a typical pattern both for percentage of correct responses and for percentage of items omitted. The results further indicate that examinees show a noticeably greater tendency to omit

Table 8. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section II: Results for High and Low Fifths on Three Recent Forms

Item Number	Item Type	Correct		Omitted		Not Reached		Not Attempted	
		High Fifth	Low Fifth	High Fifth	Low Fifth	High Fifth	Low Fifth	High Fifth	Low Fifth
1	Antonyms	95.1	51.8	0.3	4.0	0.0	0.0	0.3	4.0
2	"	97.8	31.3	0.1	10.6	0.0	0.0	0.1	10.6
3	"	98.5	21.3	0.5	18.8	0.0	0.0	0.5	18.8
4	"	87.1	38.4	2.3	10.7	0.0	0.0	2.3	10.7
5	"	86.3	17.9	2.3	16.5	0.0	0.0	2.3	16.5
6	"	86.7	4.8	2.9	26.7	0.0	0.0	2.9	26.7
7	"	38.6	−1.7	11.4	27.1	0.0	0.0	11.4	27.1
8	"	44.1	−2.8	12.7	23.1	0.0	0.0	12.7	23.1
9	"	12.1	−3.2	22.7	25.9	0.0	0.0	22.7	25.9
10	"	16.9	−7.8	20.5	19.9	0.0	0.0	20.5	19.9
11	Sentence Completion	96.6	54.2	0.1	1.4	0.0	0.0	0.1	1.4
12	"	96.3	34.3	0.2	4.4	0.0	0.0	0.2	4.4
13	"	87.5	31.4	0.8	6.7	0.0	0.0	0.8	6.7
14	"	76.7	7.2	1.3	7.1	0.0	0.0	1.3	7.1
15	"	47.0	−8.5	6.7	10.7	0.0	0.0	6.7	10.7
16	Analogies	96.8	54.9	0.0	2.3	0.0	0.0	0.0	2.3
17	"	90.9	54.0	0.4	5.5	0.0	0.0	0.4	5.5
18	"	94.8	47.6	0.2	2.6	0.0	0.0	0.2	2.6
19	"	89.8	34.5	1.3	5.4	0.0	0.0	1.3	5.4
20	"	84.1	12.7	1.7	16.2	0.0	0.0	1.7	16.2
21	"	85.2	21.0	1.2	11.8	0.0	0.0	1.2	11.8
22	"	72.5	2.4	5.1	32.2	0.0	0.0	5.1	32.2
23	"	36.2	−0.8	14.6	18.5	0.0	0.0	14.6	18.5
24	"	39.3	−0.7	25.3	36.0	0.0	0.1	25.3	36.1
25	"	22.1	−6.9	20.4	21.7	0.0	0.1	20.4	21.8
26	Reading Comprehension	69.9	7.5	1.4	6.8	0.1	0.2	1.5	7.0
27	"	70.1	11.5	1.2	7.7	0.1	0.3	1.3	8.0
28	"	51.9	−5.5	1.2	11.3	0.1	0.5	1.3	11.7
29	"	87.0	29.4	1.8	7.9	0.1	1.4	1.9	9.3
30	"	72.2	14.1	1.0	7.3	0.1	2.0	1.1	9.3
31	"	68.7	16.6	1.1	8.3	0.1	3.1	1.2	11.4
32	"	67.4	7.7	0.8	5.2	0.5	5.5	1.3	10.8
33	"	63.3	11.1	1.0	6.3	0.6	6.7	1.6	13.0
34	"	55.7	−1.1	2.5	6.7	0.8	9.8	3.2	16.5
35	"	68.9	6.0	3.7	11.4	1.0	12.6	4.7	24.0
36	"	65.0	26.7	3.7	7.9	2.2	19.8	5.9	27.7
37	"	34.4	−5.3	6.7	8.2	3.2	27.0	9.9	35.1
38	"	67.0	1.0	2.3	5.1	4.2	30.6	6.5	35.7
39	"	61.5	1.7	1.6	3.2	6.9	34.2	8.5	37.4
40	"	36.3	−2.1	0.0	0.0	12.7	40.0	12.7	40.0

Table 9. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for Four Black Groups Combined and Four White Groups Combined

Item Number	Item Type	Correct		Omitted		Not Reached		Not Attempted	
		Black	White	Black	White	Black	White	Black	White
1–15	Antonyms	27.6	39.3	15.0	15.3	0.0	0.0	15.0	15.3
16–25	Sentence Completion	25.9	39.9	5.1	4.4	0.0	0.0	5.1	4.4
26–35	Analogies	25.2	40.4	13.3	13.0	0.4	0.1	13.8	13.1
36–45	Reading Comprehension	27.3	44.2	2.9	2.8	16.7	8.8	19.6	11.6
1–10	Antonyms	25.4	37.6	15.3	15.3	0.0	0.0	15.3	15.3
11–25	Reading Comprehension	24.9	39.0	8.1	6.7	0.9	0.2	8.9	6.8
26–30	Sentence Completion	22.2	38.3	7.0	6.8	6.5	1.7	13.5	8.5
31–40	Analogies	17.4	34.3	6.6	7.5	21.6	10.9	28.2	18.4

Table 10. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Segment of the SAT-Verbal Sections: Results for Four Black Groups Separately and Four White Groups Separately

Item Number	Item Type	Group	Correct		Omitted		Not Reached		Not Attempted	
			Black	White	Black	White	Black	White	Black	White
1–15	Antonyms	WSA4	27.1	45.0	16.0	15.0	0.0	0.0	16.0	15.0
		WSA5	22.2	35.3	15.6	16.9	0.0	0.0	15.6	16.9
		XSA2-M	32.7	35.7	13.4	14.2	0.0	0.0	13.4	14.2
		XSA2-F	34.5	39.8	13.3	14.0	0.0	0.0	13.3	14.0
16–25	Sentence Completion	WSA4	31.6	51.5	5.9	4.2	0.0	0.0	5.9	4.2
		WSA5	20.2	37.5	5.0	4.0	0.0	0.1	5.0	4.0
		XSA2-M	26.8	31.4	4.4	5.3	0.1	0.0	4.5	5.3
		XSA2-F	24.9	30.1	4.1	4.6	0.0	0.0	4.2	4.6
26–35	Analogies	WSA4	26.2	46.5	17.5	16.6	0.2	0.1	17.7	16.6
		WSA5	21.3	38.3	12.1	10.4	0.6	0.2	12.7	10.6
		XSA2-M	27.2	34.2	11.0	12.3	0.6	0.1	11.7	12.4
		XSA2-F	28.8	38.2	9.8	11.8	0.4	0.0	10.2	11.8
36–45	Reading Comprehension	WSA4	32.2	54.5	3.5	2.9	14.6	6.7	18.1	9.6
		WSA5	25.1	46.9	2.9	2.5	16.9	8.5	19.9	11.0
		XSA2-M	22.9	29.9	2.4	2.8	22.6	12.8	25.0	15.6
		XSA2-F	26.1	32.4	2.2	3.0	14.6	9.8	16.8	12.8
1–10	Antonyms	WSA4	25.6	43.4	17.9	15.1	0.0	0.0	17.9	15.1
		WSA5	21.0	34.4	15.0	16.0	0.0	0.0	15.0	16.0
		XSA2-M	31.3	36.3	12.4	14.2	0.0	0.0	12.4	14.2
		XSA2-F	27.9	33.9	13.7	15.7	0.0	0.0	13.7	15.7
11–25	Reading Comprehension	WSA4	26.1	47.2	7.3	4.6	0.7	0.0	8.0	4.6
		WSA5	29.6	48.0	6.8	4.7	0.9	0.2	7.6	4.9
		XSA2-M	17.8	20.0	10.6	11.8	1.2	0.3	11.8	12.1
		XSA2-F	19.9	23.7	9.8	9.8	0.8	0.2	10.6	10.0
26–30	Sentence Completion	WSA4	20.3	40.6	10.5	9.1	5.7	1.3	16.2	10.4
		WSA5	23.4	45.0	4.7	4.7	6.5	1.8	11.2	6.5
		XSA2-M	22.9	29.7	5.6	6.6	8.8	2.0	14.4	8.6
		XSA2-F	22.6	28.5	5.8	6.8	5.7	2.1	11.5	8.9
31–40	Analogies	WSA4	21.7	39.9	7.2	8.9	21.5	9.3	28.7	18.3
		WSA5	12.1	37.0	7.1	5.9	21.7	10.4	28.9	16.3
		XSA2-M	16.9	24.4	5.0	7.4	25.8	12.8	30.7	20.2
		XSA2-F	19.7	27.7	5.8	8.0	17.6	13.0	23.5	21.0

Table 11. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section I: Results for Four Black Groups Combined and Four White Groups Combined

Item Number	Item Type	Correct		Omitted		Not Reached		Not Attempted	
		Black	White	Black	White	Black	White	Black	White
1	Antonyms	67.4	87.7	4.3	2.0	0.0	0.0	4.3	2.0
2	"	67.0	85.0	2.4	1.4	0.0	0.0	2.4	1.4
3	"	60.0	80.0	5.1	2.6	0.0	0.0	5.1	2.6
4	"	53.2	70.9	3.5	1.9	0.0	0.0	3.5	1.9
5	"	45.5	67.9	10.1	7.8	0.0	0.0	10.1	7.8
6	"	38.7	59.2	9.5	5.7	0.0	0.0	9.5	5.7
7	"	26.3	46.2	9.7	7.4	0.0	0.0	9.7	7.4
8	"	23.0	32.8	17.6	16.7	0.0	0.0	17.6	16.7
9	"	12.8	23.5	20.3	21.5	0.0	0.0	20.3	21.5
10	"	12.4	17.2	21.1	25.5	0.0	0.0	21.1	25.5
11	"	4.9	8.1	19.0	20.2	0.0	0.0	19.0	20.2
12	"	7.8	8.2	22.7	24.7	0.0	0.0	22.7	24.7
13	"	2.8	6.0	26.1	32.0	0.0	0.0	26.1	32.0
14	"	1.0	2.8	27.5	31.6	0.0	0.0	27.5	31.6
15	"	−6.9	−5.4	25.5	28.7	0.0	0.0	25.5	28.7
16	Sentence Completion	58.0	76.9	1.0	0.9	0.0	0.0	1.0	0.9
17	"	52.6	74.4	1.5	0.8	0.0	0.0	1.5	0.8
18	"	41.5	64.1	2.7	1.3	0.0	0.0	2.7	1.3
19	"	66.3	81.7	1.4	1.0	0.0	0.0	1.4	1.0
20	"	43.5	57.6	3.0	2.0	0.0	0.0	3.0	2.0
21	"	16.6	31.5	4.1	3.2	0.0	0.0	4.1	3.2
22	"	5.9	18.5	10.5	9.7	0.0	0.0	10.5	9.8
23	"	−2.8	6.5	8.7	9.3	0.1	0.0	8.8	9.3
24	"	−11.7	−4.1	4.9	3.8	0.1	0.0	4.9	3.8
25	"	−10.9	−8.2	12.7	11.8	0.1	0.0	12.8	11.9
26	Analogies	70.4	86.3	0.9	0.3	0.1	0.0	1.0	0.4
27	"	61.3	82.7	1.3	0.7	0.1	0.0	1.5	0.8
28	"	51.1	69.9	3.7	3.8	0.1	0.0	3.9	3.8
29	"	28.8	53.1	13.1	10.3	0.2	0.0	13.3	10.3
30	"	15.8	37.1	15.8	13.1	0.2	0.1	16.0	13.1
31	"	18.2	37.4	14.3	11.0	0.4	0.1	14.6	11.0
32	"	6.0	11.8	25.2	27.0	0.5	0.1	25.7	27.1
33	"	4.5	17.9	17.4	17.6	0.6	0.2	18.0	17.8
34	"	−0.5	7.5	25.2	28.8	0.9	0.2	26.1	29.0
35	"	−4.1	−0.2	16.2	17.3	1.2	0.4	17.4	17.7
36	Reading Comprehension	33.0	45.6	2.1	1.6	2.8	0.8	4.9	2.4
37	"	39.2	50.4	4.3	4.1	4.1	1.4	8.5	5.5
38	"	37.4	52.8	2.6	2.4	5.3	1.7	7.8	4.1
39	"	32.9	47.3	2.3	2.3	6.8	2.2	9.2	4.5
40	"	24.1	42.2	3.1	3.5	9.2	3.4	12.3	6.9
41	"	27.2	45.0	2.5	2.5	19.5	9.1	21.9	11.6
42	"	19.1	37.0	5.3	4.7	24.4	12.6	29.7	17.3
43	"	19.5	40.7	3.8	3.1	27.7	15.8	31.5	18.9
44	"	23.8	43.8	2.8	3.7	31.2	18.9	34.0	22.6
45	"	16.7	37.0	0.0	0.0	36.2	22.1	36.2	22.1

Table 12. Percentage of Correct, Omitted, Not Reached, and Not Attempted Items for Each Item of SAT-Verbal Section II: Results for Four Black Groups Combined and Four White Groups Combined

Item Number	Item Type	Correct		Omitted		Not Reached		Not Attempted	
		Black	White	Black	White	Black	White	Black	White
1	Antonyms	60.0	81.3	3.7	2.0	0.0	0.0	3.7	2.0
2	"	63.0	81 2	6.0	3.5	0.0	0.0	6.0	3.5
3	"	56.0	73.5	3.8	2.9	0.0	0.0	3.8	2.9
4	"	47.0	71.5	7.3	4.5	0.0	0.0	7.3	4.5
5	"	22.6	38.3	16.8	14.1	0.0	0.0	16.8	14.1
6	"	7.6	18.9	18.5	18.6	0.0	0.0	18.5	18.6
7	"	2.4	6.1	20.9	22.1	0.0	0.0	20.9	22.1
8	"	3.2	7.3	25.6	29.4	0.0	0.0	25.6	29.4
9	"	−1.1	3.2	26.2	30.0	0.0	0.0	26.3	30.0
10	"	−6.4	−5.1	24.3	26.2	0.0	0.0	24.3	26.2
11	Reading Comprehension	19.1	42.3	6.5	5.8	0.0	0.0	6.5	5.8
12	"	37.2	50.6	3.2	3.3	0.0	0.0	3.2	3.3
13	"	33.4	43.4	4.1	4.3	0.0	0.0	4.1	4.3
14	"	28.7	42.7	4.6	5.1	0.0	0.0	4.6	5.1
15	"	20.6	38.4	3.6	3.2	0.1	0.0	3.6	3.2
16	"	47.2	57.8	2.9	2.1	0.1	0.0	3.0	2.1
17	"	10.7	19.3	7.4	7.2	0.1	0.0	7.6	7.2
18	"	15.3	28.5	7.3	6.1	0.2	0.1	7.4	6.2
19	"	33.4	46.0	5.3	4.1	0.2	0.1	5.4	4.2
20	"	42.0	57.5	5.5	3.5	0.3	0.1	5.8	3.5
21	"	7.2	22.1	9.9	7.6	1.2	0.2	11.1	7.8
22	"	33.0	50.9	11.7	8.6	1.9	0.3	13.5	8.9
23	"	9.2	25.5	16.8	13.5	2.6	0.5	19.4	14.0
24	"	15.7	25.3	16.7	13.4	2.8	0.6	19.5	14.0
25	"	20.4	34.7	15.6	12.2	3.4	0.6	19.0	12.8
26	Sentence Completion	49.7	74.9	1.8	1.0	4.3	1.0	6.1	2.0
27	"	43.5	61.8	3.1	2.6	5.3	1.2	8.4	3.7
28	"	19.9	39.6	9.4	10.6	6.4	1.6	15.8	12.3
29	"	4.9	17.0	11.5	11.1	7.7	2.1	19.2	13.3
30	"	−7.3	−2.0	9.0	8.8	8.8	2.6	17.8	11.4
31	Analogies	56.2	73.6	3.8	2.0	10.5	3.4	14.3	5.4
32	"	32.6	53.2	6.0	5.9	12.0	4.2	18.0	10.2
33	"	29.1	51.2	4.9	3.0	13.8	4.8	18.7	7.9
34	"	12.8	37.5	10.0	9.9	16.1	6.0	26.2	15.9
35	"	7.9	18.4	10.3	15.4	18.0	7.3	28.3	22.7
36	"	0.3	13.4	6.6	8.0	20.5	8.4	27.0	16.4
37	"	13.8	34.5	9.8	13.2	24.4	11.0	34.2	24.1
38	"	13.3	31.1	7.1	9.6	27.7	14.4	34.8	24.0
39	"	2.2	15.8	7.1	8.2	32.8	18.6	39.9	26.8
40	"	5.3	14.6	0.0	0.0	40.6	30.5	40.6	30.5

analogies and antonyms items, on which the vocabulary subscore is based, than to omit reading comprehension and sentence completion items, on which the reading subscore is based. This pattern is apparent for examinees in the high fifth and in the low fifth as well as for the total group and for black and white examinees. To the extent that these results arise because examinees may adopt varying strategies in allocating their time to different item types, they complicate the interpretation of findings on the relative validity of the reading and vocabulary subscores and of scores based on the different item types.

RESULTS: EQUATED DELTAS AND BISERIAL CORRELATIONS

Cross-tabulations of equated deltas and biserial correlations for each item type are shown in Tables 13 through 17. Tables 13 and 14 present results for three recent forms (3CSA5, 3CSA6, 3CSA7); Tables 15 through 17 show results for each of the three forms separately.

Some indications of the contribution of different item types to fulfilling test specifications can be obtained by counting the number of items of each type having a difficulty level above a particular value. For example, Table 13 shows that 44 of the 254 items shown in the table have an equated delta of 15.0 or higher. Of these, 20 are antonyms, 12 are analogies, 7 are reading comprehension, and 5 are sentence completion items. Thus, the item types represented in the vocabulary subscore contribute more heavily to supplying these relatively difficult items. Of the 44 items, 25 items also have a biserial correlation of .30 or higher. Of these 25 items, 8 are antonyms, 7 are analogies, 5 are reading comprehension, and 5 are sentence completion items. Although antonyms and analogies contribute three-fifths of these items, it is clear that all four item types are supplying some items in this category.

The foregoing comparisons do not take account of the number of items of each type included in the three forms. Table 14 shows the percentage of items of each type having various values of equated deltas and biserial correlations. When the four item types are compared on the basis of the percentage having equated deltas of 15.0 or higher, the following results are obtained: antonyms, 26.7 percent; analogies, 23.3 percent; reading comprehension, 9.3 percent; and sentence completion 11.1 percent. In the three forms studied, the two item types included in the vocabulary score provide a substantially greater proportion of the items with equated deltas of 15.0 or higher. However, when the further requirement that items have a biserial correlation of .30 or higher is considered, the percentages for each item type are as follows: antonyms, 10.7 percent; analogies, 11.9 percent; reading comprehension, 6.7 percent; and sentence completion, 11.1 percent. These results indicate that when account is taken of the number of items of each type included in the SAT-verbal sections, only reading comprehension is noticeably lower in the percentage of items having the designated level of equated deltas and biserial correlations. This finding would be expected because there is less flexibility in choosing reading comprehension items than is true for the other item types.

Tables 15 through 17 show cross-tabulations of equated deltas and biserial correlations for each of the three forms separately. The detailed results differ somewhat from form to form, as would be expected. However, they suggest that antonyms, analogies, and sentence completion are about equally good sources of relatively difficult items with reasonably high biserial correlations. Although reading comprehension provides some items in this category, its contribution is less, relatively, than that of the other item types.

Table 18 shows the distribution of equated deltas for the three forms, separately and combined. The results indicate that the sentence completion and analogies items were relatively easy, on the average, and reading comprehension items were relatively difficult, on the average. With respect to the standard deviation of deltas, the reading comprehension items in these forms showed a relatively small mean value (2.69) as compared with the other three item types, for which standard deviations ranged from 3.37 to 3.65.

Table 19, which shows the distribution of biserial correlations, indicates a noticeably higher mean (.52) for sentence completion than for the other three item types, for which means range from .46 to .48. None of the 45 sentence completion items and only two of the 75 reading comprehension items had biserial correlations below .30. For antonyms and analogies, the corresponding figures are 14 of 75 and 8 of 59, respectively. This result suggests that antonyms and analogies items may be more likely to be measuring abilities that differ from those measured by the total score or that the items may be ambiguous.

RESULTS: TRUE-SCORE CORRELATIONS

Beginning with the test analysis of the October 1979 form of the SAT, test analyses have regularly included true-score correlations between scores on each of the four item types. Table 20 shows the true-score correlations of each of the six pairs of scores for 20 recent forms of the SAT-verbal sections, and the rank of each correlation within the set of six coefficients for each form. The results are based on reasonably large samples (ranging in size from 1,345 to 2,360). Reliability coefficients were calculated using Dressel's (1940) adaptation of KR 20. Although there is of course substantial variability in the correlations for a given pair, the rank ordering of most pairs shows a fairly consistent pattern.

The fact that the true-score correlations depend not only on the observed correlation but also on reliability estimates for each of the two item scores has two consequences that deserve mention. First, factors such as

Table 13. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Frequency of Each Item Type for Three Recent Forms Combined. Note: Biserial and Delta not Computed for One Analogies Item

		Equated Delta														
Biserial Correlation	*Item Type*	*5.9 Down*	*6.0–6.9*	*7.0–7.9*	*8.0–8.9*	*9.0–9.9*	*10.0–10.9*	*11.0–11.9*	*12.0–12.9*	*13.0–13.9*	*14.0–14.9*	*15.0–15.9*	*16.0–16.9*	*17.0–17.9*	*18.0–18.9*	*Total*
.80–.89	Antonyms															1
	Analogies															
	Reading Comprehension															
	Sentence Completion															
.70–.79	Antonyms				1	1										2
	Analogies															
	Reading Comprehension															
	Sentence Completion	1														1
.60–.69	Antonyms		5	1	2	3	3	2	2				1			19
	Analogies		1			1			1	1						4
	Reading Comprehension			1	1		1						1			4
	Sentence Completion		1		3	1			3			1				9
.50–.59	Antonyms	2	2	3		4	2	2	1		1	1				18
	Analogies	3	2	5	4	2	3	1	1	1	2			1		25
	Reading Comprehension		1	1	1	6	3	4	5	1	1	2				25
	Sentence Completion	4	2	2	1			1	1	3	2	1				17
.40–.49	Antonyms	1	1	1	1		1	1		3	3	2	2			16
	Analogies		3		1	1	3		1	1		1	1			12
	Reading Comprehension		3	1	5	2	2	3	3	4	6					29
	Sentence Completion	1		1	2	1	2	2		2	1	2				14
.30–.39	Antonyms						2				1	1	1			5
	Analogies			1	1						4	2	1	1		10
	Reading Comprehension				1		1	2	2	5	2	1	1			15
	Sentence Completion				1						2	1				4
.20–.29	Antonyms					1					1	2	4	1		9
	Analogies			1							1	4				6
	Reading Comprehension														1	1
	Sentence Completion															
.10–.19	Antonyms											4		1		5
	Analogies										1	1				2
	Reading Comprehension													1		1
	Sentence Completion															
Total	Antonyms	3	9	5	4	9	8	5	3	3	6	10	8	2		75
	Analogies	3	6	7	6	4	6	1	3	3	8	8	2	2		59
	Reading Comprehension		4	3	8	8	7	9	10	10	9	3	2	1	1	75
	Sentence Completion	6	3	3	7	2	2	3	4	5	5	5				45

Table 14. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Percentage of Each Item Type for Three Recent Forms Combined. Note: Biserial and Delta not Computed for One Analogies Item

		Equated Delta														
Biserial Correlation	*Item Type*	*5.9 Down*	*6.0–6.9*	*7.0–7.9*	*8.0–8.9*	*9.0–9.9*	*10.0–10.9*	*11.0–11.9*	*12.0–12.9*	*13.0–13.9*	*14.0–14.9*	*15.0–15.9*	*16.0–16.9*	*17.0–17.9*	*18.0–18.9*	*Total*
.80–.89	Antonyms		1.3													1.3
	Analogies															
	Reading Comprehension															
	Sentence Completion															
.70–.79	Antonyms				1.3	1.3										2.7
	Analogies															
	Reading Comprehension															
	Sentence Completion	2.2														2.2
.60–.69	Antonyms		6.7	1.3	2.7	4.0	4.0	2.7	2.7				1.3			25.3
	Analogies		1.7			1.7			1.7	1.7						6.8
	Reading Comprehension			1.3	1.3		1.3						1.3			5.3
	Sentence Completion		2.2		6.7	2.2			6.7			2.2				20.0
.50–.59	Antonyms	2.7	2.7	4.0		5.3	2.7	2.7	1.3		1.3	1.3				24.0
	Analogies	5.1	3.4	8.5	6.8	3.4	5.1	1.7	1.7	1.7	3.4			1.7		42.4
	Reading Comprehension		1.3	1.3	1.3	8.0	4.0	5.3	6.7	1.3	1.3	2.7				33.3
	Sentence Completion	8.9	4.4	4.4	2.2			2.2	2.2	6.7	4.4	2.2				37.8
.40–.49	Antonyms	1.3	1.3	1.3	1.3		1.3	1.3		4.0	4.0	2.7	2.7			2[illegible]
	Analogies		5.1		1.7	1.7	5.1		1.7	1.7		1.7	1.7			20.3
	Reading Comprehension		4.0	1.3	6.7	2.7	2.7	4.0	4.0	5.3	8.0					38.7
	Sentence Completion	2.2		2.2	4.4	2.2	4.4	4.4		4.4	2.2	4.4				31.1
.30–.39	Antonyms						2.7				1.3	1.3	1.3			6.7
	Analogies			1.7	1.7						6.8	3.4	1.7	1.7		16.9
	Reading Comprehension				1.3		1.3	2.7	2.7	6.7	2.7	1.3	1.3			20.0
	Sentence Completion				2.2						4.4	2.2				8.9
.20–.29	Antonyms					1.3					1.3	2.7	5.3	1.3		12.0
	Analogies			1.7							1.7	6.8				10.2
	Reading Comprehension														1.3	1.3
	Sentence Completion															
.10–.19	Antonyms											5.3		1.3		6.7
	Analogies										1.7	1.7				3.4
	Reading Comprehension													1.3		1.3
	Sentence Completion															
Total	Antonyms	4.0	12.0	6.7	5.3	12.0	10.7	6.7	4.0	4.0	8.0	13.3	10.7	2.7		100.0
	Analogies	5.1	10.2	11.9	10.2	6.8	10.2	1.7	5.1	5.1	13.6	13.6	3.4	3.4		100.0
	Reading Comprehension		5.3	4.0	10.7	10.7	9.3	12.0	13.3	13.3	12.0	4.0	2.7	1.3	1.3	100.0
	Sentence Completion	13.3	6.7	6.7	15.6	4.4	4.4	6.7	8.9	11.1	11.1	11.1				100.0

Table 15. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Percentage of Each Item Type for Form 3CSA5

		Equated Delta													
Biserial Correlation	*Item Type*	*5.9 Down*	*6.0–6.9*	*7.0–7.9*	*8.0–8.9*	*9.0–9.9*	*10.0–10.9*	*11.0–11.9*	*12.0–12.9*	*13.0–13.9*	*14.0–14.9*	*15.0–15.9*	*16.0–16.9*	*17.0–17.9*	*18.0–18.9*	*Total*
.80–.89	Antonyms		4.0													4.0
	Analogies															
	Reading Comprehension															
	Sentence Completion															
.70–.79	Antonyms				4.0											4.0
	Analogies															
	Reading Comprehension															
	Sentence Completion															
.60–.69	Antonyms		4.0						8.0							12.0
	Analogies															
	Reading Comprehension				4.0								4.0			8.0
	Sentence Completion				13.3				6.7							20.0
.50–.59	Antonyms	4.0				4.0	4.0	4.0								16.0
	Analogies	5.0	5.0	5.0	5.0	10.0	5.0	5.0	5.0		10.0			5.0		60.0
	Reading Comprehension		4.0			8.0	4.0	8.0	8.0	4.0		4.0				40.0
	Sentence Completion	6.7						6.7	6.7		13.3					33.3
.40–.49	Antonyms	4.0	4.0	4.0	4.0					8.0	4.0		4.0			22.0
	Analogies		10.0				5.0									15.0
	Reading Comprehension		4.0	4.0	4.0	8.0	4.0		4.0	12.0	4.0					44.0
	Sentence Completion			6.7	6.7		6.7	6.7		6.7	6.7	6.7				46.7
.30–.39	Antonyms						4.0					4.0				8.0
	Analogies												5.0	5.0		10.0
	Reading Comprehension								4.0							4.0
	Sentence Completion															
.20–.29	Antonyms					4.0					4.0	4.0	12.0			24.0
	Analogies			5.0								5.0				10.0
	Reading Comprehension														4.0	4.0
	Sentence Completion															
.10–.19	Antonyms															
	Analogies											.50				5.0
	Reading Comprehension															
	Sentence Completion															
Total	Antonyms	8.0	12.0	4.0	8.0	8.0	8.0	4.0	8.0	8.0	8.0	8.0	16.0			100.0
	Analogies	5.0	15.0	10.0	5.0	10.0	10.0	5.0	5.0		10.0	10.0	5.0	10.0		100.0
	Reading Comprehension		8.0	4.0	8.0	16.0	8.0	8.0	16.0	16.0	4.0	4.0	4.0		4.0	100.0
	Sentence Completion	6.7		6.7	20.0		6.7	13.3	13.3	6.7	20.0	6.7				100.0

Table 16. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Percentage of Each Item Type for Form 3CSA6. Note: Biserial and Delta not Computed for One Analogies Item

		Equated Delta													
Biserial Correlation	*Item Type*	*5.9 Down*	*6.0–6.9*	*7.0–7.9*	*8.0–8.9*	*9.0–9.9*	*10.0–10.9*	*11.0–11.9*	*12.0–12.9*	*13.0–13.9*	*14.0–14.9*	*15.0–15.9*	*16.0–16.9*	*17.0–17.9*	*18.0–18.9*	*Total*
.80–.89	Antonyms															
	Analogies															
	Reading Comprehension															
	Sentence Completion															
.70–.79	Antonyms					4.0										4.0
	Analogies															
	Reading Comprehension															
	Sentence Completion	6.7														6.7
.60–.69	Antonyms		4.0	4.0	4.0	8.0	4.0	4.0					4.0			32.0
	Analogies								5.3	5.3						10.5
	Reading Comprehension			4.0			4.0									8.0
	Sentence Completion								6.7			6.7				13.3
.50–.59	Antonyms		8.0	8.0		4.0		4.0			4.0	4.0				32.0
	Analogies	5.3	5.3	21.1	5.3											36.8
	Reading Comprehension				4.0	8.0	4.0	8.0	4.0			4.0				32.0
	Sentence Completion	13.3		13.3	6.7					20.0						53.3
.40–.49	Antonyms						4.0				4.0	8.0	4.0			20.0
	Analogies		5.3		5.3	5.3							5.3			21.1
	Reading Comprehension		4.0		4.0			4.0	8.0		8.0					28.0
	Sentence Completion				6.7	6.7				6.7						20.0
.30–.39	Antonyms						4.0									4.0
	Analogies										10.5	10.5				21.1
	Reading Comprehension				4.0				4.0	16.0	4.0					28.0
	Sentence Completion										6.7					6.7
.20–.29	Antonyms											10.5				10.5
	Analogies															
	Reading Comprehension															
	Sentence Completion															
.10–.19	Antonyms											4.0		4.0		8.0
	Analogies															
	Reading Comprehension													4.0		4.0
	Sentence Completion															
Total	Antonyms		12.0	12.0	4.0	16.0	12.0	8.0			8.0	16.0	8.0	4.0		100.0
	Analogies	5.3	10.5	21.1	10.5	5.3			5.3	5.3	10.5	21.1	5.3			100.0
	Reading Comprehension		4.0	4.0	12.0	8.0	8.0	12.0	16.0	16.0	12.0	4.0		4.0		100.0
	Sentence Completion	20.0		13.3	13.3	6.7			6.7	26.7	6.7	6.7				100.0

Table 17. Cross-Tabulations of Biserial Correlation Coefficients and Equated Deltas for Four SAT-Verbal Item Types: Percentage of Each Item Type for Form 3CSA7

		Equated Delta														
Biserial Correlation	*Item Type*	*5.9 Down*	*6.0–6.9*	*7.0–7.9*	*8.0–8.9*	*9.0–9.9*	*10.0–10.9*	*11.0–11.9*	*12.0–12.9*	*13.0–13.9*	*14.0–14.9*	*15.0–15.9*	*16.0–16.9*	*17.0–17.9*	*18.0–18.9*	*Total*
.80–.89	Antonyms															
	Analogies															
	Reading Comprehension															
	Sentence Completion															
.70–.79	Antonyms															
	Analogies															
	Reading Comprehension															
	Sentence Completion															
.60–.69	Antonyms		12.0		4.0	4.0	8.0	4.0								32.0
	Analogies		5.0			5.0										10.0
	Reading Comprehension															
	Sentence Completion		6.7		6.7	6.7			6.7							26.7
.50–.59	Antonyms	4.0		4.0		8.0	4.0		4.0							24.0
	Analogies	5.0			10.0		10.0			5.0						30.0
	Reading Comprehension			4.0		8.0	4.0		8.0		4.0					28.0
	Sentence Completion	6.7	13.3									6.7				26.7
.40–.49	Antonyms							4.0		4.0	4.0					12.0
	Analogies						10.0		5.0	5.0		5.0				25.0
	Reading Comprehension		4.0		12.0		4.0	8.0		4.0	12.0					44.0
	Sentence Completion	6.7					6.7	6.7				6.7				26.7
.30–.39	Antonyms										4.0		4.0			8.0
	Analogies			5.0	5.0						10.0					20.0
	Reading Comprehension						4.0	8.0		4.0	4.0	4.0	4.0			28.0
	Sentence Completion				6.7						6.7	6.7				20.0
.20–.29	Antonyms											4.0	4.0	4.0		12.0
	Analogies										5.0	5.0				10.0
	Reading Comprehension															
	Sentence Completion															
.10–.19	Antonyms											12.0				12.0
	Analogies										5.0					5.0
	Reading Comprehension															
	Sentence Completion															
Total	Antonyms	4.0	12.0	4.0	4.0	12.0	12.0	8.0	4.0	4.0	8.0	16.0	8.0	4.0		100.0
	Analogies	5.0	5.0	5.0	15.0	5.0	20.0		5.0	10.0	20.0	10.0				100.0
	Reading Comprehension		4.0	4.0	12.0	8.0	12.0	16.0	8.0	8.0	20.0	4.0	4.0			100.0
	Sentence Completion	13.3	20.0		13.3	6.7	6.7	6.7	6.7		6.7	20.0				100.0

Table 18. Frequency Distribution of Equated Deltas by Item Type of Three Recent Forms of SAT-Verbal Sections. Note: Equated Delta not Computed for One Analogies Item

Equated Delta	Antonyms				Analogies				Reading Comprehension				Sentence Completion			
	3CSA5	3CSA6	3CSA7	Total	3CSA5	3CSA6	3CSA7	Total	3CSA5	3CSA6	3CSA7	Total	3CSA5	3CSA6	3CSA7	Total
18.0–18.9	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
17.0–17.9	0	1	1	2	2	0	0	2	0	1	0	1	0	0	0	0
16.0–16.9	4	2	2	8	1	1	0	2	1	0	1	2	0	0	0	0
15.0–15.9	2	4	4	10	2	4	2	8	1	1	1	3	1	1	3	5
14.0–14.9	2	2	2	6	2	2	4	8	1	3	5	9	3	1	1	5
13.0–13.9	2	0	1	3	0	1	2	3	4	4	2	10	1	4	0	5
12.0–12.9	2	0	1	3	1	1	1	3	4	4	2	10	2	1	1	4
11.0–11.9	1	2	2	5	1	0	0	1	2	3	4	9	2	0	1	3
10.0–10.9	2	3	3	8	2	0	4	6	2	2	3	7	1	0	1	2
9.0– 9.9	2	4	3	9	2	1	1	4	4	2	2	8	0	1	1	2
8.0– 8.9	2	1	1	4	1	2	3	6	2	3	3	8	3	2	2	7
7.0– 7.9	1	3	1	5	2	4	1	7	1	1	1	3	1	2	0	3
6.0– 6.9	3	3	3	9	3	2	1	6	2	1	1	4	0	0	3	3
5.9 Down	2	0	1	3	1	1	1	3	0	0	0	0	1	3	2	6
Total	25	25	25	75	20	19	20	59	25	25	25	75	15	15	15	45
Mean	11.28	11.19	11.60	11.36	11.07	10.87	11.28	11.08	11.52	11.72	11.66	11.63	11.25	10.31	10.16	10.58
S.D.	3.82	3.57	3.55	3.65	4.00	3.76	3.07	3.63	2.97	2.55	2.52	2.69	2.86	3.48	3.61	3.37

Table 19. Frequency Distribution of Biserial Correlation by Item Type of Three Recent Forms of SAT-Verbal Sections. Note: Biserial not Computed for One Analogies Item

Biserial Correlation	Antonyms				Analogies				Reading Comprehension				Sentence Completion			
	3CSA5	3CSA6	3CSA7	Total	3CSA5	3CSA6	3CSA7	Total	3CSA5	3CSA6	3CSA7	Total	3CSA5	3CSA6	3CSA7	Total
.90–.94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.85–.89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.80–.84	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
.75–.79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.70–.74	1	1	0	2	0	0	0	0	0	0	0	0	0	1	0	1
.65–.69	1	5	3	9	0	0	2	2	2	1	0	3	1	1	0	2
.60–.64	2	3	5	10	0	2	0	2	0	1	0	1	2	1	4	7
.55–.59	4	5	3	12	4	5	2	11	6	3	2	11	3	5	2	10
.50–.54	0	3	3	6	8	2	4	14	4	5	5	14	2	3	2	7
.45–.49	6	2	2	10	2	3	3	8	5	4	5	14	4	1	0	5
.40–.44	2	3	1	6	1	1	2	4	6	3	6	15	3	2	4	9
.35–.39	2	1	1	4	2	2	4	8	1	5	5	11	0	0	2	2
.30–.34	0	0	1	1	0	2	0	2	0	2	2	4	0	1	1	2
.25–.29	5	0	1	6	1	2	2	5	1	0	0	1	0	0	0	0
.20–.24	1	0	2	3	1	0	0	1	0	0	0	0	0	0	0	0
.15–.19	0	1	1	2	1	0	1	2	0	1	0	1	0	0	0	0
.10–.14	0	1	2	3	0	0	0	0	0	0	0	0	0	0	0	0
.05–.09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.00–.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	25	25	25	75	20	19	20	59	25	25	25	75	15	15	15	45
Mean	0.47	0.52	0.46	0.48	0.47	0.47	0.45	0.46	0.49	0.45	0.44	0.46	0.52	0.54	0.49	0.52
S.D.	0.15	0.14	0.18	0.16	0.11	0.12	0.12	0.12	0.09	0.10	0.07	0.09	0.08	0.09	0.10	0.09

Table 20. True-Score Correlations of SAT-Verbal Item Types for 20 SAT Forms

Administration	*Sentence Completion vs. Reading Comprehension*		*Sentence Completion vs. Antonyms*		*Sentence Completion vs. Analogies*		*Reading Comprehension vs. Antonyms*		*Reading Comprehension vs. Analogies*		*Antonyms vs. Analogies*	
	Corr	*Rank*	*Corr*	*Rank*	*Corr*	*Rank*	*Corr*	*Rank*	*Corr*	*Rank*	*Corr*	*Rank*
October 1979	.855	5	.947	1	.835	6	.875	4	.883	3	.903	3
November 1979	.953	3	.973	2	.939	4	.904	5	.902	6	.974	1
December 1979	.959	1	.952	2	.910	3	.901	4	.865	6	.888	5
January 1980	.956	1	.942	2	.909	3	.909	4	.874	6	.907	5
March 1980	.949	3	.967	1	.943	4	.897	6	.898	5	.960	2
May 1980	.946	3	.950	2	.952	1	.909	5	.866	6	.912	4
June 1980	.901	3	.970	1	.885	5	.903	2	.848	6	.889	4
October 1980	.940	3	.975	1	.945	2	.871	6	.876	5	.916	4
October 1980	.942	2	.967	1	.910	5	.912	4	.863	6	.918	3
December 1980	.913	4	.926	3	.936	1.5	.883	5	.875	6	.936	1.5
January 1981	.913	4	.974	1	.934	2	.866	5	.850	6	.925	3
April 1981	.925	3.5	.953	1	.925	3.5	.881	5	.874	6	.941	2
May 1981	.947	1	.917	2	.908	5	.908	4	.884	6	.911	3
May 1981	.962	2	.990	1	.944	3	.907	3	.907	6	.922	4
June 1981	.943	3	.952	2	.966	1	.891	5	.867	6	.909	4
November 1981	.949	2	.984	1	.939	3	.873	6	.887	5	.921	4
December 1981	.889	4	.960	1	.941	2	.873	5	.856	6	.935	3
January 1982	.912	5	.956	1	.955	2	.886	6	.915	4	.941	3
January 1982	.963	3	.980	1	.965	2	.897	5	.886	6	.925	4
March 1982	.942	2	.974	1	.909	5	.911	4	.855	6	.931	3
Median	.942	2	.964	1	.938	3	.897	5	.874	6	.922	4

speededness that may introduce bias into reliability estimates will also affect the true-score correlation. Second, because the same reliability estimate is used for all true-score correlations of a particular item score, any systematic error in estimating the reliability will also produce a systematic error in the true-score correlations for that score.

The following table shows the median correlation of each part with each other part for the 20 forms:

	Antonyms	Analogies	Reading Comprehension	Sentence Completion
Antonyms	—	.922	.897	.964
Analogies	.922	—	.874	.938
Reading Comprehension	.897	.874	—	.942
Sentence Completion	.964	.938	.942	—

Certain features of these results deserve comment, as follows:

1. Reading comprehension has a noticeably higher correlation with sentence completion (.942) than with antonyms (.897) or with analogies (.874). This result is consistent with the grouping of reading comprehension and sentence completion in defining the reading score.
2. Sentence completion has relatively high correlations (.938 to .964) with the other three item type scores. In view of its high correlation with antonyms (.964), it is somewhat surprising that its correlation with reading comprehension (.942) is noticeably higher than the correlation of antonyms with reading comprehension (.897). It should also be noted that the correlation between analogies and sentence completion (.938) is slightly higher than the correlation between analogies and antonyms (.922).

Although these results suggest that the present allocation of item types to the two SAT-verbal subscores is reasonable, it appears that sentence completion is a good measure of vocabulary as well as of reading. There would be, accordingly, some logical basis for including sentence completion in the vocabulary score rather than in the reading score. This format would avoid the inclusion of passage-type items with other item types within the same time limit and would provide vocabulary and reading subscores each based on a separately timed subtest. However, this approach would make it very difficult to obtain a satisfactory distribution of item difficulties for the reading subscore and would result in some reduction in reliability for the reading subscore. An alternative plan calling for the same revised test format would continue to include sentence completion items in the reading score.

Under the alternative plan, the introduction of a separately timed reading comprehension section would be less likely to affect comparability of scores for the old and new formats. The evaluation of the relative merits of those plans is difficult, however, without evidence on the validity of the four item types that constitute the SAT-verbal sections.

Finally, true-score correlations of vocabulary and reading were summarized for the same 20 forms. The following distribution was observed:

Correlation	Frequency
.955–.959	1
.950–.954	1
.945–.949	3
.940–.944	8
.935–.939	6
.930–.934	1

The median correlation based on this distribution is .941. This result indicates that the abilities measured by the reading subscore and the vocabulary subscore are very similar. Although this high true-score correlation is desirable as evidence that the SAT-verbal score is measuring a reasonably homogeneous set of abilities, the high correlation also raises some question about the usefulness of reporting separate subscores for such similar tests.

SUMMARY AND CONCLUSIONS

Summaries of item analysis data clearly indicate that the percentage of examinees omitting or not reaching an item tends to increase substantially for the more difficult items that are placed at the end of each segment. This tendency appears to be particularly marked for the antonyms and analogies segments. The pattern of results lends credibility to the hypothesis that examinees may differ in the relative proportion of their time and effort that they allocate to reading passages and items as compared with the other three item types. Although the data do not provide a basis for estimating the extent of this variation, they do suggest that the organization of the test into segments and the apparent difficulty of items affect examinees' test-taking strategies. Because the four item types are not administered with separate time limits, variations in examinee strategies may well affect both reliability and validity coefficients for item scores and for the vocabulary and reading subscores.

Interpretation of the results of this survey is complicated by the fact that the item types were not administered in separately timed test sections and by the fact that the four item types were not matched with respect to level and dispersion of item difficulties in the test forms studied. Because item difficulty as judged by the examinees can be expected to affect their decision on attempting an item, and because relatively difficult items having high biserial coefficients are relatively rare, the observed differences in results for the four item types may have arisen in part by the way the forms studied were assembled. In these forms, it appears that reading comprehension provides a smaller proportion of items that have both relatively high difficulty and relatively high biserial correlations than do the other three item types. Antonyms and analogies appear to be good sources of relatively difficult items, but also have a somewhat larger proportion of items that have a relatively low biserial correlation than the other two item types.

Summaries of true-score correlations between each pair of item types show that sentence completion correlates noticeably higher, on the average, with reading comprehension than do antonyms and analogies. However, sentence completion shows quite high correlations with antonyms and analogies, indicating that these three item types are measuring quite similar abilities. Accordingly, insofar as the true-score correlations are concerned, the possibility of basing the reading subscore on a 30-minute section composed only of reading comprehension items and basing the vocabulary score on a 30-minute section composed of the other three item types might deserve consideration. An alternative plan would call for the same format but would continue to base the reading score on both reading comprehension and sentence completion items.

The Relation of SAT Reading and Vocabulary Scores to Measures of High School Performance

ABSTRACT

Relationships of reading and vocabulary subscores, used alone and also in combination with other variables, to high school rank, to Achievement Test scores, to self-reported course grade in various subjects, and to essay and objective scores on the English Composition Test were studied. For most measures of high school performance, reading yielded higher correlations than vocabulary but lower correlations than SAT-verbal scores. Multiple correlations would be little affected by revising the relative weight of reading and vocabulary in defining SAT-verbal scores when scores are used in combination with other ability measures.

INTRODUCTION

Beginning with the 1974–75 testing year, two verbal subscores have been reported for SAT-verbal sections: vocabulary, based on analogies and antonym items, and reading, based on reading comprehension and sentence completion items. In preparing a chapter on the predictive validity of the SAT for the forthcoming edition of the *College Board Technical Manual*, Ramist[1] observed that in a group of 96 colleges the reading comprehension subscores had an average validity of .370 against freshman GPA as compared with an average validity of .320 for vocabulary. Indeed, in 42 of these colleges, reading had a higher validity than either SAT-verbal total score or the vocabulary subscore. In a subsequent memorandum, Ramist[2] reported that when 110 colleges were stratified on the basis of their SAT-verbal mean score into 8 groups, reading had a higher validity against freshman GPA in each of the eight groups. For the 110 colleges, the average validity coefficient was .376 for SAT-verbal, .373 for reading, and .321 for vocabulary.

Ramist's finding that the average validity coefficient for the SAT-verbal score is only .003 higher than the average coefficient for reading alone indicates that the vocabulary items in the SAT-verbal sections are contributing remarkably little to correlation with college grades. His findings point to a need for further information about the abilities measured by reading and vocabulary items.

The main purpose of this study is to learn more about similarities and differences between reading comprehension and vocabulary in the abilities that each measures.

One way to learn more about what SAT-verbal subscores are measuring is to correlate them with familiar measures of high school performance. For example, correlations of reading and vocabulary with the College Board Achievement Tests show how much the relative size of the correlations for the two subscores varies from one academic field to another. In addition, multiple correlation analyses using measures of high school performance as criteria should provide a useful preliminary estimate of how effective optimal weighting would be if similar analyses were made using college grades as criteria.

One incidental outcome of this study deserves mention. Unlike the correlations between SAT scores and high school rank obtained in validity studies, which necessarily are based on enrolled students, the correlations in this study are based on substantially all SAT-takers in the 1981 cohort. Correlations based on SAT-takers probably provide a more useful measure of the degree of similarity between SAT scores and high school rank in the abilities that they measure than correlations based on enrolled students, mainly because relatively few SAT-takers who have low scores *and* low high school rank are admitted to selective colleges. To the extent that high school rank and SAT scores are used jointly in selection, as is usually the case, the correlation between these variables tend to be lower (and in some cases much lower) in the selected group than in the applicant group for that college.

In this study, each of the 15 Achievement Tests, self-reported course grades in the subject area corresponding to each Achievement Test, and self-reported high school rank were studied. Data for these analyses were obtained from the College Board 1981 Senior File, which includes the latest score on each test taken and data from the Student Descriptive Questionnaire for members of the cohort completing high school in 1981. In addition, results of an unpublished study by Petersen and Livingston[3] relating reading and vocabulary subscores, SAT-verbal scores, and self-reported high school rank to essay and objective scores on the December 1977 English Composition Test were made

1. L. Ramist, "Validity of the SAT-Verbal Subscores," Memorandum, February 12, 1981.
2. L. Ramist, "Further Investigation of the Validity of SAT-Verbal Subscores," Memorandum, July 7, 1981.
3. N.S. Petersen and S.A. Livingston, "English Composition Test with Essay: A Descriptive Study of the Relationship between Essay and Objective Scores by Ethnic Group and Sex," Unpublished Report, July 1982.

Table 1. Intercorrelations of SAT Scores and High School Rank (Based on 849,175 Examinees in 1981 Senior File)

	Reading	*Vocabulary*	*SAT-V*	*SAT-M*	*TSWE*	*High School Rank*
Reading	—	.814	.950	.663	.743	.462
Vocabulary	.814	—	.952	.647	.753	.439
SAT-V	.950	.952	—	.688	.786	.473
SAT-M	.663	.647	.688	—	.638	.523
TSWE	.743	.753	.786	.638	—	.462
High School Rank	.462	.439	.473	.523	.462	—

available for this study. The Petersen and Livingston study included separate analyses for white, black, Mexican American, and Oriental groups. A summary of these results is included in this review.

These studies of the relationship of SAT scores to various measures of high school performance throw considerable light on the nature of the abilities measured by the reading and vocabulary subtests, and should be taken into account in evaluating possible changes in the composition of the SAT-verbal sections.

METHOD

For the analyses of the 1981 senior file, 16 tables of intercorrelations were computed. One table included all students who had complete data on SAT scores, including the two verbal subscores, Test of Standard Written English (TSWE), and high school rank. Each of the remaining 15 tables included all students who had complete data on the six variables included in the initial table, a particular Achievement Test, and self-reported grades (A through D) on the course area most relevant to the Achievement Test. Multiple regression analyses were calculated using high school rank, each of the 15 Achievement Tests, and the course grade appropriate to each Achievement Test as dependent variables. The multiple regression analyses were done in sets of four in order to permit comparisons of (1) reading, (2) vocabulary, (3) SAT-verbal, and (4) both reading comprehension and vocabulary when each of the four was used in conjunction with other variables. For high school rank, the SAT-mathematical score was the additional variable. For each Achievement Test, SAT-mathematical score and high school rank were the additional variables. For course grades, two sets of analyses were performed. In one, only the SAT-mathematical score was used as the additional variable. In the other, both SAT-mathematical score and the designated Achievement Test score served as additional variables.

Results for the study of essay and objective scores on the English Composition Test were obtained from the draft report of the Petersen and Livingston study. Multiple correlations were calculated based on results for the total group (using correlation coefficients correct to four decimal places) with essay and objective grades and high school rank as dependent variables.

RESULTS

As shown in Table 1, correlation with high school rank is higher for reading (.462) than for vocabulary (.439) but lower than for the SAT-verbal score (.473). Multiple correlation analysis results, shown in Table 2, however, indicate that when used along with the SAT-mathematical score, the advantage of reading over vocabulary is slight (.545 versus .540). Using vocabulary and reading in combination with the SAT-mathematical score yields a multiple of .546 (only .001 higher than the coefficient

Table 2. Multiple Correlation with High School Rank of Combinations of SAT Scores (Based on 849,175 Examinees in 1981 Senior File)

Combination of Scores	*Multiple Correlation*
Reading, SAT-M	.545
Vocabulary, SAT-M	.540
SAT-V, SAT-M	.546
Reading, Vocabulary, SAT-M	.546

obtained without vocabulary), the same coefficient that is obtained when the SAT-verbal score is used in combination with the SAT-mathematical score.

Table 3 shows the correlations of SAT scores (and subscores) and high school rank with Achievement Test scores. For Russian, as it happened, both reading and vocabulary have negative correlation (−.247 for reading versus −.329 for vocabulary). For the other 14 Achievement Tests, reading has a higher positive correlation with Achievement Test scores for every test except Latin. For Latin, the difference is very small (.497 for vocabulary versus .494 for reading.)

For the 14 comparisons excluding Russian, the median difference is .038, with reading showing the higher correlation. Comparing SAT-verbal correlations with those for reading alone aids in evaluating the contribution of vocabulary items to SAT-verbal scores. Based on the 14 tests excluding Russian, the SAT-verbal score has a higher correlation than reading in 9 of the 14 comparisons. The median difference is .013, indicating that the vocabulary items made only a small contribution to the correlations.

Table 4 shows the multiple correlations when SAT-mathematical scores and high school rank are used along with the three verbal scores. Comparisons may be made first for results when reading is used along with the other two variables with results when vocabulary is used along with the same two variables. For the four tests in English and social studies, three of the four comparisons show a higher correlation for the combination involving reading, with differences ranging from .012 to .021. As it happens, for English Composition, the combination involving vocabulary is higher by .002. For the two mathematical tests, the differences are trivial, as would be expected because verbal scores receive very little weight compared to SAT-mathematical scores. Among the six language tests, vocabulary has a negative regression weight for Hebrew and both vocabulary and reading have a negative weight for Russian. For the other four languages, the combination involving reading is higher for German, by .035, and for Spanish, by .020, but the combination involving vocabulary is higher for French, by .006, and for Latin, by .010. For the three science tests, the combination involving reading is higher in all three comparisons, with differences ranging from .006 to .012. Excluding Russian and the two mathematics tests, the combination involving reading yields a higher multiple correlation in 9 of 12 comparisons, with a median difference of .012.

Comparisons of the multiple correlations obtained for the combination involving SAT-verbal scores with the combination in which reading and vocabulary are weighted optimally reveal that for only three tests (German, Hebrew, and Spanish) does the four-variable combination yield a correlation more than .002 higher than the three-variable combination. On the whole, it appears that reading is somewhat more closely related to Achievement Test scores than is vocabulary and that SAT-verbal scores are slightly more closely related to Achievement Test scores than is reading. When scores are used in combination with SAT-mathematical scores and high school rank, the differences between reading and vocabulary remain, but are reduced in size. Moreover, when optimal weights are given to reading and vocabulary as well as to SAT-mathematical scores and high school rank, multiple correlations are only trivially larger than those obtained when SAT-verbal scores are combined with SAT-mathematical scores and high school rank.

Table 3. Correlation of SAT Scores and High School Rank with Achievement Test Scores

Achievement Test	*N*	*Reading*	*Vocabulary*	*SAT-V*	*SAT-M*	*High School Rank*
English Composition	149,452	.755	.748	.792	.561	.415
Literature	12,737	.820	.795	.845	.530	.399
American History and Social Studies	44,761	.730	.705	.756	.525	.344
European History and World Cultures	2,573	.701	.670	.718	.436	.273
Mathematics, Level I	118,665	.507	.466	.514	.823	.444
Mathematics, Level II	31,744	.510	.453	.508	.806	.366
French	17,832	.454	.450	.477	.398	.284
German	3,802	.313	.260	.303	.215	.171
Hebrew	303	.295	.159	.238	.305	.248
Latin	1,615	.494	.497	.521	.481	.338
Russian	244	−.247	−.329	.307	.127	.109
Spanish	20,276	.222	.181	.212	.168	.194
Biology	30,980	.698	.663	.718	.672	.416
Chemistry	28,721	.580	.536	.589	.650	.354
Physics	13,209	.577	.510	.572	.671	.297

Table 4. Multiple Correlation of Various Combinations of SAT Scores and High School Rank with Achievement Test Scores

Achievement Test	*N*	*Reading, SAT-M, High School Rank*	*Vocabulary, SAT-M, High School Rank*	*SAT-V, SAT-M, High School Rank*	*Reading, Vocabulary, SAT-M, High School Rank*
English Composition	149,452	.775	.777	.804	.805
Literature	12,737	.824	.806	.847	.848
American History and Social Studies	44,761	.745	.733	.767	.767
European History and World Cultures	2,573	.705[a]	.684[a]	.722[a]	.723[a]
Mathematics, Level I	118,665	.830	.830	.830	.831
Mathematics, Level II	31,744	.814	.813	.814	.815
French	17,832	.490	.496	.505	.505
German	3,802	.321	.286	.313	.321
Hebrew	303	.381	.331[b]	333	.381[b]
Latin	1,615	.558	.568	.572	.572
Russian	244	.376[c]	.426[c]	.425[c]	.427[c]
Spanish	20,276	.251	.231	.244	.251
Biology	30,980	.763	.756	.774	.774
Chemistry	28,721	.702	.696	.706	.706
Physics	13,209	.720[a]	.708[a]	.719[a]	.721[a]

[a]Negative regression weights for high school rank.
[b]Negative regression weights for vocabulary.
[c]Negative regression weights for reading comprehension, vocabulary, and SAT-V.

Table 5. Correlation of Test Scores with Course Grade for Students Taking Each Achievement Test

Course Grade	*Achievement Test Taken*	*N*	*Reading*	*Vocabulary*	*SAT-V*	*SAT-M*	*Achievement Test Score*
English	English Composition	149,452	.315	.306	.327	.222	.344
English	Literature	12,737	.339	.336	.352	.228	.358
Social Studies	American History and Social Studies	44,761	.250	.230	.254	.212	.275
Social Studies	European History and World Cultures	2,573	.294	.286	.305	.219	.274
Mathematics	Mathematics, Level I	118,665	.186	.164	.185	.378	.409
Mathematics	Mathematics, Level II	31,744	.161	.136	.157	.317	.327
Foreign Language	French	17,832	.216	.210	.225	.209	.259
Foreign Language	German	3,802	.197	.190	.206	.181	.177
Foreign Language	Hebrew	303	.261	.240	.266	.227	.130
Foreign Language	Latin	1,615	.234	.247	.252	.222	.270
Foreign Language	Russian	244	.102	.074	.088	.173	.155
Foreign Language	Spanish	20,276	.198	.196	.206	.208	.219
Biological Sciences	Biology	30,980	.322	.302	.329	.319	.360
Physical Sciences	Chemistry	28,721	.274	.247	.275	.349	.332
Physical Sciences	Physics	13,209	.275	.253	.278	.342	.300

Correlations of SAT scores and Achievement Test scores with course grades in the subject area most appropriate for each Achievement Test are shown in Table 5. Interpretation of results for course grades should take account of the fact that students were asked to report their latest year-end or midyear grade in each subject matter area. Thus, the course grade is based on one or, at most, two semesters of work. In some instances, the grade may be in a

different subject from that covered in the Achievement Test. Although those conditions may affect the size of the correlations, it seems unlikely that they would affect the comparisons with which this study is concerned.

As shown in Table 5, reading yields a higher correlation with grades than does vocabulary in 14 of 15 comparisons, the exception being Latin. For the four English and social studies tests, differences range from .003 (Literature) to .020 (American History and Social Studies) with a median of .008. For foreign languages, differences range from −.013 (Latin) to .028 (Russian) with a median of .006. For mathematics and sciences, differences range from .022 (Mathematics, Level I) to .027 (Chemistry) with a median of .022.

When correlations of SAT-verbal scores with the course grades are compared with the corresponding correlations for reading, the SAT-verbal correlation is higher in 12 of 15 comparisons, the exceptions arising for the two mathematics groups and for the group taking the Russian test. The median of the 15 differences is .008, and the largest difference is .018, for the group taking the Latin test. SAT-verbal scores yield correlations that are slightly larger than those for reading, a result that is consistent with the findings for the Achievement Tests.

Table 6 shows results obtained when reading, vocabulary, and SAT-verbal scores are used along with SAT-mathematical scores. The first two columns provide evidence that the combination involving reading yields a higher correlation than the combination involving vocabulary in nine comparisons, the largest difference being .009 (for Hebrew). In six comparisons, the combination including vocabulary yields the higher correlation, the largest difference being .013 (for Latin). When the coefficient for the vocabulary combination is subtracted from the coefficient for the reading combination, the median of the 15 differences is .002.

When an optimally weighted combination of reading, vocabulary, and SAT-mathematical scores is compared with an optimally weighted combination of SAT-verbal and SAT-mathematical scores, eight of the differences in the multiples are .000, and in only two instances does the four-variable combination exceed the three-variable combination by as much as .002. For two groups, German and European History and World Cultures, the two-variable combination is higher (by .001) than the three-variable combination. (This puzzling result may have resulted from the fact that SAT-verbal scores are based on the sum of the *raw* scores on reading and vocabulary but the multiple correlations involving reading and vocabulary are based on *scaled* scores.) The results indicate that changing the relative weights for vocabulary and reading scores would have only a trivial effect on the multiple correlation of SAT-verbal and SAT-mathematical scores with course grades.

Table 7 shows the multiple correlations obtained when both the appropriate Achievement Test and SAT-mathematical scores are used in combination with SAT scores. The

Table 6. Multiple Correlation with Course Grade of Combination of SAT Scores for Students Taking Each Achievement Test

Course Grade	*Achievement Test Taken*	*N*	*Reading, SAT-M*	*Vocabulary, SAT-M*	*SAT-V, SAT-M*	*Reading, Vocabulary, SAT-M*
English	English Composition	149,452	.319	.314	.330	.330
English	Literature	12,737	.341	.339	.353	.354
Social Studies	American History and Social Studies	44,761	.265	.255	.267	.268
Social Studies	European History and World Cultures	2,573	.304	.302	.313	.312
Mathematics	Mathematics, Level I	118,665	.379[a]	.380[b]	.380[b]	.380[b]
Mathematics	Mathematics, Level II	31,744	.317[a]	.318[b]	.318[b]	.318[b]
Foreign Language	French	17,832	.240	.241	.245	.245
Foreign Language	German	3,802	.215	.216	.221	.220
Foreign Language	Hebrew	303	.279	.270	.282	.282
Foreign Language	Latin	1,615	.256	.269	.268	.270
Foreign Language	Russian	244	.175	.174	.174	.176[b]
Foreign Language	Spanish	20,276	.227	.228	.231	.231
Biological Sciences	Biology	30,980	.356	.351	.360	.360
Physical Sciences	Chemistry	28,721	.362	.359	.362	.363
Physical Sciences	Physics	13,209	.360	.358	.361	.361

[a]Negative regression weights for reading comprehension, vocabulary, and SAT-V.
[b]Negative regression weights for vocabulary.

Table 7. Multiple Correlation with Course Grade of Various Combinations of SAT Scores and Achievement Test Scores for Students Taking Each Achievement Test

Course Grade	*Achievement Test Taken*	*N*	*Reading, SAT-M, Achievement Test*	*Vocabulary, SAT-M, Achievement Test*	*SAT-V, SAT-M, Achievement Test*	*Reading, Vocabulary, SAT-M, Achievement Test*
English	English Composition	149,452	.354	.352	.356	.356
English	Literature	12,737	.367	.368	.370	.370
Social Studies	American History and Social Studies	44,761	.290	.288	.290	.290
Social Studies	European History and World Cultures	2,573	.316	.316	.321	.321
Mathematics	Mathematics, Level I	118,665	.418*	.418*	.419*	.419*
Mathematics	Mathematics, Level II	31,744	.340*	.341*	.341*	.341*
Foreign Language	French	17,832	.290	.290	.292	.292
Foreign Language	German	3,802	.244	.248	.250	.249
Foreign Language	Hebrew	303	.282	.277	.287	.286
Foreign Language	Latin	1,615	.300	.306	.304	.306
Foreign Language	Russian	244	.233	.232	.233	.234
Foreign Language	Spanish	20,276	.286	.289	.289	.290
Biological Sciences	Biology	30,980	.382	.380	.382	.383
Physical Sciences	Chemistry	28,721	.379	.378	.379	.379
Physical Sciences	Physics	13,209	.365	.364	.366	.366

*Negative regression weights for reading comprehension, vocabulary, and SAT-V.

combination including reading yields a higher multiple with course grades than the combination including vocabulary in 7 comparisons, and the combination including vocabulary yields the higher multiple in 5 comparisons. There are three ties. When optimally weighted reading and vocabulary subscores are used in conjunction with SAT-mathematical scores and other appropriate Achievement Test scores, the multiple correlation with grades exceeds the corresponding multiple based on SAT-verbal, SAT-mathematical, and Achievement Test scores in only 4 of the 15 comparisons. In only one instance (Latin) is the increase as much as .002. In 2 comparisons (German and Hebrew), the four-variable combination yields a multiple that is .001 lower than the multiple for the three-variable combination.

Table 8. Mean Score on Reading and Vocabulary for Students Taking Each Test

Test	*N*	*Reading Mean*	*Vocabulary Mean*	*Difference*
Russian	244	53.33	51.28	+2.05
Physics	13,209	54.18	52.67	+1.51
Mathematics, Level II	31,744	55.49	54.11	+1.38
Hebrew	303	53.54	52.42	+1.12
Chemistry	28,721	53.72	52.75	+ .97
European History and World Cultures	2,573	55.76	54.97	+ .79
German	3,802	53.90	53.26	+ .64
American History and Social Studies	44,761	50.72	50.29	+ .43
English Composition	149,452	50.97	50.60	+ .37
Literature	12,737	52.14	51.82	+ .32
Mathematics, Level I	118,665	49.64	49.44	+ .20
Biology	30,980	51.93	51.76	+ .17
SAT	849,175	42.88	42.77	+ .11
Spanish	20,276	49.66	49.68	− .02
French	17,832	53.63	53.79	− .16
Latin	1,615	55.32	55.48	− .16

In summary, reading used alone shows a somewhat higher correlation with course grade than does vocabulary used alone, and the SAT-verbal score correlates only slightly higher than does reading with course grade. This result suggests that the correlations of the SAT-verbal score, used alone, with course grade would be enhanced if the relative weight given to reading items were increased and that there would be only a slight loss in validity if vocabulary items were dropped. On the other hand, the multiple correlation results indicate that there would be virtually no advantage in modifying the existing relative weights for reading and vocabulary in determining SAT-verbal scores when these scores are used along with SAT-mathematical scores or with SAT-mathematical and the appropriate Achievement Test scores.

The analysis of data for the students taking each Achievement Test provided data on the mean reading and mean vocabulary subscore for each group. Comparison of those means for the 15 groups should throw some light on the relation, if any, between the difference in means and other characteristics of the group. Table 8 shows the means

for the students taking each Achievement Test and for the group taking the SAT. Groups are listed in order of difference between reading and vocabulary means. Students taking Russian, Physics, Mathematics Level II, Hebrew, and Chemistry show noticeably higher means for reading. Twelve of the 15 groups of Achievement Test takers show higher means for reading than for vocabulary, with differences ranging from .17 to 2.05. The reading mean for all SAT takers in the sample is also higher than the vocabulary mean, but the difference is only .11. Only three groups (Latin, French, and Spanish) show mean differences in the reverse direction. The presence of nontrivial, systematic differences in the means for different examinee groups can be taken as evidence that reading and vocabulary, although they measure similar abilities, are not strictly parallel.

Table 9. Correlation of Reading, Vocabulary, and SAT-V with Measures of Ability in English and with High School Rank

	Group[a]	*Reading*	*Vocabulary*	*SAT-V*
ECT Essay	T	.46	.47	.49
	W	.44	.44	.47
	B	.46	.47	.50
	M-A	.45	.44	.48
	O-EBL	.51	.52	.55
ECT Objective	T	.70	.72	.75
	W	.69	.70	.74
	B	.70	.73	.76
	M-A	.67	.70	.73
	O-EBL	.72	.72	.76
ECT Composite	T	.71	.72	.76
	W	.69	.71	.74
	B	.70	.73	.76
	M-A	.68	.69	.73
	O-EBL	.73	.74	.78
TSWE	T	.66	.67	.71
	W	.64	.65	.68
	B	.69	.71	.74
	M-A	.67	.67	.71
	O-EBL	.69	.71	.74
High School Rank	T	.39	.38	.41
	W	.40	.39	.42
	B	.27	.26	.28
	M-A	.27	.28	.29
	O-EBL	.38	.37	.40
English Grade	T	.32	.31	.34
	W	.32	.32	.34
	B	.23	.24	.25
	M-A	.25	.28	.28
	O-EBL	.32	.29	.32

[a]Groups were defined as follows:
T = Total group of 55,673 students who took the English Composition Test in December 1977
W = White students (N = 48,920)
B = Black students (N = 2,137)
M-A = Mexican-American students (N = 649)
O-EBL = Oriental students whose best language is English (1,935).

Table 9 is based on results obtained as part of the Petersen and Livingston study. This study provides evidence on the correlation between the three scores based on the SAT-verbal sections and six measures of high school performance: English Composition Test essay, objective, and composite scores, TSWE scores, self-reported English grade, and self-reported high school rank. Analyses for the total group and four subgroups (white, black, and Mexican American students, and Oriental students whose best language is English). All members of the sample took the English Test in December 1977.

For the essay scores on the ECT, vocabulary had a slightly higher correlation (by .01) than reading for the total group and for two of the four subgroups. For Mexican American students, reading has a higher correlation (by .01) than vocabulary. The SAT-verbal score has a higher correlation than either subscore in all five comparisons. For objective and composite scores on the ECT, vocabulary has higher correlations (by amounts ranging from .01 to .03) than reading in 9 of 10 comparisons. For the same 10 comparisons, the SAT-verbal score has a higher correlation than vocabulary (by .03 or .04) in all instances. On the whole, results for the Test of Standard Written English are similar to those for objective scores on the ECT, although there are detailed differences between the two sets of results.

High school rank shows a higher correlation (by .01) for reading than for vocabulary in the total group and three of the subgroups. For Mexican American students, vocabulary is higher (by .01). The SAT-verbal score has a higher correlation than either subscore in all five comparisons.

Results for English grade are somewhat more variable from one subgroup to another than is true for the other measures. Reading has a higher correlation than vocabulary for the total group (by .01) and for Orientals whose best

Table 10. Correlation with English Composition Test Essay and Objective Scores with SAT Scores (Based on 55,673 Students Who Took the English Composition Test in December 1977)

SAT Score(s)	*ECT Essay Score*	*ECT Objective Score*	*High School Rank*
Reading	.462	.703	.390
Vocabulary	.468	.716	.377
SAT-Verbal	.493	.752	.406
Reading, Vocabulary	.492[a]	.753[a]	.407[a]

[a]Multiple correlations.

language is English (by .03). For white students, the two subscores have equal correlations. Vocabulary has the higher correlation for black students (by .01) and for Mexican American students (by .03). The SAT-verbal score has a correlation that is higher than either subscore in three comparisons but equal to the higher subscore correlation in the other two comparisons.

Table 10 shows the correlations of reading, vocabulary, and SAT-verbal scores and the multiple correlation of the two subscores with three measures of high school performance: ECT essay, ECT objective, and high school rank. These results offer no reason to believe that shifting the relative weight of reading and vocabulary items would result in noticeably higher correlations of SAT-verbal scores with these three measures of high school performance.

SUMMARY AND CONCLUSIONS

This study is concerned with exploring the relationship of reading and vocabulary subscores, singly and in combination with each other and with other variables, to measures of high school performance.

Reading tends to correlate more highly than does vocabulary with Achievement Test scores, high school rank, and course grades, based on data for the 1981 high school senior cohort. This result is consistent with Ramist's findings for the prediction of college grades. The Petersen and Livingston results, however, showed higher correlations for vocabulary than for reading with essay, objective, and composition scores on the English Composition Test, and with TSWE scores for their total group of students who took the December 1977 ECT. Only for high school rank and English grade was the correlation higher for reading.

In general, SAT-verbal scores tend to yield higher correlations with measures of high school performance than do either of the two subscores, although the differences are relatively small for the 1981 Senior File data. The results seem consistent with the hypothesis that although reading and vocabulary measure very similar abilities, reading gives slightly more weight than vocabulary to academic achievement. On the very important topic of the relation to English Composition Test scores and high school English grade, reading and vocabulary show about equal correlations when both the study based on the 1981 Senior File and the Petersen and Livingston study are taken into account.

The extensive multiple correlation results obtained in this study suggest that differences between reading and vocabulary are markedly diminished when other ability measures are used in combination with them. Finally, the results strongly indicate that assigning optimal weights to reading and vocabulary in defining SAT-verbal scores would have virtually no effect on *multiple* correlations including SAT-verbal scores with the measures of high school performance used in this study.

On the whole, the results indicate that insofar as correlations with high school rank, with Achievement Test scores, and with high school grades in particular subject areas are concerned, reading tends to yield higher correlations than does vocabulary.

It is possible to estimate the effect on the correlations between SAT-verbal scores and other variables that would be obtained if the SAT-verbal sections were composed entirely of reading items. In order to calculate this effect, it is necessary to estimate: (1) the reliability of the present test and (2) the proportion of the testing time now devoted to reading items (Gulliksen 1950, pp. 88–90). For the reliability of the present reading test, the median reliability of five new forms introduced between October 1980 and April 1981—.848—was used. In order to estimate the proportion of time devoted to reading items in current test forms, estimates of the time per item for each of the four item types[4] and the number of items of each type were used. It turns out that approximately 41 minutes is allocated to reading items, and 19 minutes is allocated to vocabulary items in current forms of the SAT-verbal sections. Thus, if all of the SAT-verbal sections were composed of reading items, the length of the present reading test would be multiplied by 1.463 (60/41). Using a formula provided by Gulliksen, lengthening the reading test by a factor of 1.463 would be expected to multiply its correlation with other variables by

$$\sqrt{\frac{1.463}{1 + (.463)(.848)}} = 1.025.$$

Although this estimate is approximate, particularly because the reading test is not separately timed, it probably yields a reasonable estimate of the effect of devoting all of the SAT-verbal sections to reading on its correlation with other variables. For example, using data from Table 1, the estimated correlation of a 60-minute reading test with high school rank would be .462 x 1.025, which equals .474, a value which is very slightly higher than the value of .473 for the present SAT-verbal score. Similarly, if the same estimating procedure is applied to the correlations in Table 3 of reading with the 14 Achievement Tests other than Russian, the adjusted coefficients are higher than those for SAT-verbal scores in 8 of 14 comparisons. The median difference is .003.

REFERENCES

Gulliksen, H. 1950. *Theory of Mental Tests*. New York: Wiley.

4. Pamela I. Cruise, Personal communication.

The Validity of SAT-Verbal Item Types

ABSTRACT

This study is concerned with the relative validities for predicting college freshman grades of the four item types of which the SAT-verbal sections are composed. The study design provided for obtaining the validity of the four item types both for all items in the December 1980 form of the test and for subsets of items of each type matched with respect to the number of items and difficulty level. Data are analyzed for 48 colleges, each of which had a sample of 95 or more students.

The findings of this study must be regarded as tentative, particularly because only one form of the SAT-verbal sections could be studied and because each of the matched subsets included only eight items. Results indicate that differences in relative validity of the four item types are relatively small for the test form studied, and that when matched subsets of items are compared, reading comprehension has somewhat lower validity than the other three item types.

PURPOSE

The main objective of this study is to compare the effectiveness of the four item types that constitute the SAT-verbal sections for predicting first-year college average grades. Comparisons will be based on the correlation of each item type with grades and on the contribution of each item type to prediction when used along with the other three item types and with SAT-mathematical scores, a measure of high school academic record, and Test of Standard Written English (TSWE) scores. As part of this study, comparisons will be made based on subsets of each item type composed of items attempted by at least 85 percent of students in a special analysis sample and matched with respect to the number of items and the mean and standard deviation of item difficulty indexes. Analysis of the data for matched subsets should throw some light on the effect of nonattempts and differences in difficulty distributions on the relative validity of the four item types.

BACKGROUND

Although a great many studies of the validity of SAT-verbal scores have been done, relatively little attention has been given in recent years to the relative validity of the four item types of which the SAT-verbal sections are composed.

An early study done by Olsen and Schrader (1953) was based on the item types included in one 1946 form and one 1947 form of the SAT. In this study, first-year grades were used as the criterion for the item analyses, and comparisons were based on the percentage of items of each type that exceeded the median biserial correlation for all verbal items. On the whole, the validity results were inconclusive, although the results suggested that reading had a smaller proportion of highly valid items than analogies, opposites, or completion (now called sentence completion) in the 1947 form.

A later study based on administration of specially developed tests in 1954 at five colleges classified item types on the basis of their correlations with SAT-mathematical items (Olsen and Schrader 1957). On the whole, the results suggested that no substantial increase in validity could be achieved by distinguishing between the two verbal and two mathematical item groups defined in the study. However, there is some suggestion in the results that completion and "nonoverlapping analogies" (analogies in which neither word in the matching pair is synonymous with either word in the given pair) correlated higher with grades and noticeably lower with SAT-mathematical scores than with paragraphs and overlapping analogies.

Because the Selective Service College Qualification Test was quite similar to the SAT in a number of respects, the extensive validation research conducted on this test is relevant to the evaluation of SAT-verbal item types. A study comparing short verbal (analogies and double definitions) with reading comprehension based on 13 freshman groups, 7 sophomore groups, and 6 junior groups showed no clear superiority for either item type in correlation with grades earned during the academic year when the test was given, although reading items had a higher correlation than the short verbal items in all six of the junior groups (Educational Testing Service 1953).

Although these studies were judged to be inconclusive, an unpublished study of the LSAT conducted in the early 1950s resulted in the dropping of short verbal items while substantial weight was given to reading recall in subsequent forms of the examination. Similar results were obtained for the ATGSB (now GMAT) resulting in the dropping of short verbal item types while giving substantial weight to reading recall (Pitcher 1960).

In preparing the chapter on predictive validity for the ATP technical manual, Ramist[1] observed that the reading subscore of the SAT-verbal score, based on reading comprehension and sentence completion items, had an average validity of .370 against freshman grades while the

1. L. Ramist, "Validity of the SAT-Verbal Subscores," Memorandum, February 12, 1981.

vocabulary subscore, based on analogies and antonyms items, had an average validity of .320 in a group of 96 colleges. Indeed, in 42 of the 96 colleges, reading alone had a higher validity than either total score on the SAT-verbal sections or the vocabulary test. There was only one college in which vocabulary had a higher validity than either total SAT-verbal score or reading. These results suggest that the current allocation of time to the various SAT-verbal item types is not optimal. In a subsequent memorandum, Ramist[2] reported that reading yielded consistently higher validities when colleges were stratified by the SAT-verbal mean into 8 groups. Breland[3] summarized a number of relevant findings, including the fact that in three recent forms of the SAT-verbal sections, the reading and vocabulary scores had equal reliability (.84).

As Angoff[4] has stated, confirmatory factor studies of the SAT (Rock and Werts 1979) and of the GRE (Rock, Werts, and Grandy 1981) provide relevant information about the abilities measured by verbal item types. The study of the SAT suggests that antonyms resembles reading comprehension and sentence completion more closely than it resembles analogies. The GRE factor analysis suggested that the reading comprehension items seemed to be more closely related to certain item types included in the analytical score than was true for the other three verbal item types. This result is consistent with Ramist's finding, in that the four verbal item types, despite their high intercorrelations, may differ from each other in their correlation with other variables.

When the validity findings reported by Ramist for vocabulary and reading scores are considered in the light of the factor analysis results, the need for information on the validity of the four SAT-verbal item types is clear. Such information should be helpful in planning future forms of the SAT-verbal sections.

DATA BASE

In order to obtain scores on each of the four item types for this study, a record of each examinee's responses is needed. At the time the present study was initiated, only the unedited scanning tapes for the December 1980 administration were available. Identification data common to the scanning tapes and the validity study files and used in this study included the first four letters of the last name, initials, sex, month and day of birth, and SAT-verbal, SAT-mathematical, and TSWE scores. This matching process resulted in the exclusion of some students from the validity sample who would have been retained if the edited response tapes had been available. However, requiring an exact match of scores should minimize the risk that faulty matching of data from the two data files would occur in the analysis sample. Matching of the item response tapes with the appropriate validity study data tapes yielded an initial sample of 12,031 students who had first-year average grades and were enrolled in one of 50 colleges that had 98 or more students in the sample. This sample was used to perform a special item analysis to be used in creating matching subsets of items. The sample used in the correlational analysis was limited to students having complete data on scores, high school academic record, and first-year average grade. (The measure of high school academic record chosen by the college for its validity study was used in this study.) The final sample included 48 colleges that had 95 or more students with complete data.

CONSTRUCTING THE MATCHING SUBSETS

Using standard statistical procedures, the percentage of students attempting each item and the customary index of difficulty, delta, were calculated using the initial validity sample. The matched subsets were defined so that each subset would have the same number of items and so that the

Table 1. Raw Deltas of Items Having 85 Percent or More Attempts and 8-Item Subsets of Each Type Matched on Difficulty Level

Antonyms	*Sentence Completion*	*Analogies*	*Reading Comprehension*
17.7	16.0	14.6*	15.7
15.0*	14.0*	12.0*	15.3*
12.0*	12.8*	11.9*	15.0
11.7*	12.1*	11.5*	14.9
11.7	10.9*	11.2	14.5
11.4*	10.1*	10.2*	13.8
10.8*	9.9*	10.2	13.2
10.5*	8.6*	9.9*	13.2
9.5	8.1*	9.6*	12.7
9.2*	8.1	9.0	12.0
8.4	8.0	7.6*	11.7*
8.0	6.9	7.0	10.8*
7.8*			10.7*
7.1			10.5*
			10.5*
			10.1*
			8.6*
Mean** 11.0	10.8	10.9	11.0
S.D.** 2.0	1.9	1.9	1.8

*Items selected for 8-item subsets.
**Values for 8-item subsets.

2. L. Ramist, "Further Investigation of the Validity of SAT-Verbal Subscores," Memorandum, July 7, 1981.
3. H. Breland, "Reading and Vocabulary Items in SAT," Memorandum, March 27, 1981.
4. W. H. Angoff, "Differential Validity of the Reading and Vocabulary Subscores of the SAT-Verbal," Memorandum, May 14, 1981.

means and standard deviations of item difficulties would agree closely with each other. It was decided that subsets composed of eight items would permit satisfactory matching of item difficulties. (In two instances, one of two items having equal difficulties was selected by chance for the subset.) It was also decided that only those items that were attempted by 85.0 percent or more of the students should be chosen for the matched subsets of items in order to limit the effect on the results of student decisions to omit items and of the location of items near the end of a separately timed section.

Table 1 shows the item difficulty index for each item attempted by 85.0 percent or more of the sample, the difficulty index of each item selected for the matched subset, and the mean and standard deviation of deltas for each selected subset.

ANALYSIS METHODS

Two main approaches were followed in the data analysis. The basic analysis provided validity coefficients for each predictor and multiple correlation coefficients for various combinations of predictors for each of the 48 college samples separately. The alternative approach provided for pooling data across colleges. The procedure was equivalent to expressing each variable as standard scores within each college and then combining the standard scores across colleges. It was thought that this procedure should greatly reduce the effect on the results of differences among colleges in standards of admission and grading. The coefficients for data pooled across colleges were actually calculated by determining the mean of within-college correlations. Except for the comparisons of men and women, results for each college were weighted by the number of students in the sample for that college, a procedure that gave equal weight to each student in the overall sample. For the men–women comparisons, however, this procedure would give different relative weight to a college in the sample of women from the corresponding relative weight in the sample of men. For this reason, unweighted means of the within-college correlations were computed for the men–women comparisons.

RESULTS

The main outcome of this study is evidence on the relative validity of the four item types. The study design permitted two separate comparisons of relative validity for each sample studied. One set of comparisons is based on matched subsets of items in which each item type is represented by eight items that were attempted by 85 percent or more of the validity study sample and selected so that the distribution of item difficulties would be similar for all four item types. The other set of comparisons is based on all items of that type in the December 1980 form of the SAT-verbal sections.

Six sets of comparative validities were obtained as follows:

1. *All Colleges (Median).* These validities were obtained by determining the median validity coefficient for the 48 colleges in the study, based on a separate analysis in each college. The sample included 11,320 students.
2. *All Colleges (Pooled Data).* These validities were obtained by pooling data across all colleges giving equal weight to each of the 11,320 students.
3. *Women.* These validities were obtained by pooling data for 33 colleges that had at least 50 women and 50 men in this sample, giving equal weight to each college in calculating the coefficients. The sample included 4,802 women.
4. *Men.* These validities are based on the same 33 colleges as in (3) and were obtained by using equal weights, as in the study of women. The sample size for men was also 4,802, as it happened.
5. *Higher Scoring Colleges.* These coefficients were obtained by pooling data across seven colleges that had the highest SAT-verbal mean scores. All mean scores were 480 or higher. In pooling the data, equal weight was given to each of the 1,086 students in these college samples.
6. *Lower Scoring Colleges.* These coefficients were obtained by pooling data for the seven colleges that had the lowest SAT-verbal mean scores. All mean scores were 409 or lower. In pooling the data, equal weight was given to each of the 1,458 students on these college samples.

Although great care is taken to make forms of the SAT-verbal sections parallel to each other, information on the correlation of items with grades is not available when tests are assembled. The extent to which validities of item types vary from one form to another is not known. For this reason, the fact that only one form of the SAT-verbal sections could be included in this study seriously limits the extent to which the results can be generalized beyond the particular test form studied.

Table 2 presents in summary form the comparative validities of the four item types and of reading and vocabulary subscores. (For the matched subset analyses, the subscores were the sums of scores on the appropriate matched subsets.)

For the matched subset analyses, the most noteworthy finding is the fact that reading comprehension has somewhat lower validities than the other three item types. In all six comparisons, the other three item types have a validity

equal to or (in nearly all cases) higher than reading comprehension.

Another comparison of the four item types can be made by counting the number of colleges in which each item type has the highest validity of the four and the number in which it has the lowest validity, using the results shown in Table 3. (If two item types tie for highest or lowest validity, each tie is counted as one-half.) The results for the matched subsets are as follows:

	Lowest	Highest
Antonyms	9	16
Analogies	5½	16
Reading Comprehension	15½	8½
Sentence Completion	18	7½
Total Colleges	48	48

When this method of summarizing the results is used, reading comprehension and sentence completion tend to have lower validities than analogies and antonyms. Except for sentence completion, these results are consistent with those shown in Table 2.

Table 2 also permits a comparison of validities for the matched vocabulary subscore based on analogies and antonyms items and the matched reading subscore based on reading comprehension and sentence completion items. The results indicate that vocabulary has a higher validity than reading in all six comparisons. Comparison of validities in each of the 48 colleges using data from Table 3 shows a higher validity for vocabulary in 33 colleges and a higher validity for reading in 15 colleges.

These results of the matched subsets analysis are based on only eight items for each item type and are based on only a single test form. They deserve consideration, however, because they are consistent with the possibility that differences in validity of item types in operational test forms may arise, in part, from differences in the difficulty distributions for different item types or to differences in candidate strategy in allocating their time to different item types.

Comparisons of item type validities based on all items in the operational test form can be made using the results shown in the lower portion of Table 2. Results for the four item types do not show a consistent pattern, but reading comprehension and analogies tend to have slightly higher validities than the other two item types. For the two SAT-verbal subscores, reading shows a slightly higher validity than vocabulary in both analyses that were based on all 48 colleges, but the differences are smaller than those found by Ramist.

Table 2. Correlations with First-Year Average Grades of Scores on Matched Subsets of Items and on All Items for Each Item Type

	Item Type Scores				*Subscores*	
Analysis	*Antonyms*	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Vocabulary*	*Reading*
			Scores Based on Matched Subsets			
All 48 Colleges (Median)	.18	.22	.18	.20	.26	.22
All 48 Colleges (Pooled Data)	.20	.21	.19	.19	.24	.22
Women (33 Colleges)	.22	.21	.18	.20	.25	.23
Men (33 Colleges)	.20	.22	.18	.19	.25	.22
Seven Higher Scoring Colleges	.23	.21	.18	.20	.25	.23
Seven Lower Scoring Colleges	.20	.24	.16	.21	.26	.22
Number of Items	8	8	8	8	16	16
			Scores Based on All Items			
All 48 Colleges (Median)	.23	.26	.24	.24	.26	.28
All 48 Colleges (Pooled Data)	.22	.24	.25	.21	.25	.26
Women (33 Colleges)	.23	.24	.25	.23	.26	.27
Men (33 Colleges)	.22	.25	.24	.22	.25	.26
Seven Higher Scoring Colleges	.26	.28	.26	.25	.29	.28
Seven Lower Scoring Colleges	.21	.26	.24	.23	.26	.27
Number of Items	25	20	25	15	45	40

3. Correlations of Scores Based on Matched Subsets of Items with First-Year Average Grades in Each College

Antonyms			*Analogies*			*Reading Comprehension*			*Sentence Completion*			*Vocabulary*			*Reading*			*Subset Total*			*Number of Cases*
R	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	
0.248	4.7	2.1	0.286	4.9	2.0	0.245	4.8	1.9	0.243	5.0	2.1	0.307	9.7	3.6	0.291	9.8	3.3	0.335	19.5	6.1	913
0.145	5.3	1.9	0.158	5.5	1.7	0.183	5.3	1.7	0.149	5.3	1.9	0.179	10.8	3.1	0.196	10.6	3.0	0.212	21.4	5.4	602
0.229	4.8	2.3	0.191	4.8	2.3	0.195	4.9	2.0	0.240	5.0	2.2	0.234	9.5	4.1	0.252	9.9	3.7	0.263	19.4	7.2	592
0.153	4.7	2.2	0.194	4.8	2.1	0.149	4.8	1.9	0.140	4.8	2.1	0.198	9.5	3.8	0.169	9.6	3.4	0.203	19.1	6.5	573
0.256	4.0	2.2	0.250	3.8	2.2	0.190	4.2	2.0	0.263	4.2	2.1	0.296	7.8	3.8	0.274	8.5	3.4	0.325	16.3	6.3	538
0.180	4.8	2.0	0.193	4.9	1.9	0.225	4.9	1.8	0.138	5.0	2.0	0.212	9.7	3.5	0.218	9.9	3.1	0.241	19.7	5.9	469
0.332	5.2	2.1	0.290	4.9	1.9	0.205	5.0	1.9	0.283	5.2	2.0	0.359	10.1	3.5	0.290	10.2	3.3	0.353	20.3	6.3	463
0.175	5.0	2.1	0.129	5.0	1.9	0.184	5.1	1.8	0.184	5.1	1.9	0.182	10.0	3.3	0.226	10.2	3.0	0.232	20.2	5.5	453
0.170	4.3	2.2	0.144	4.2	2.3	0.168	4.6	1.9	0.127	4.5	2.1	0.178	8.5	3.9	0.172	9.1	3.4	0.196	17.6	6.5	389
0.201	4.5	2.3	0.194	4.5	2.1	0.194	4.7	2.0	0.216	4.7	2.2	0.226	9.0	3.8	0.243	9.4	3.5	0.258	18.4	6.6	356
0.233	5.5	2.2	0.265	5.5	2.1	0.159	5.6	1.9	0.226	5.7	2.1	0.275	11.0	3.9	0.224	11.2	3.5	0.268	22.2	6.9	278
0.063	5.1	2.0	0.195	5.2	2.0	0.142	5.1	2.0	0.120	5.1	2.2	0.149	10.3	3.4	0.152	10.3	3.5	0.166	20.6	6.3	278
0.275	4.9	2.2	0.272	4.8	2.1	0.206	4.8	1.9	0.249	4.7	2.1	0.322	9.6	3.6	0.262	9.5	3.5	0.327	19.1	6.3	259
0.232	4.6	2.2	0.123	4.4	2.0	0.175	4.8	1.8	0.263	4.7	2.1	0.219	9.0	3.5	0.258	9.5	3.4	0.263	18.5	6.2	235
0.087	4.8	1.8	0.005	4.8	1.8	0.062	4.9	1.7	0.010	4.7	1.8	0.054	9.6	3.2	0.044	9.6	2.8	0.057	19.1	5.1	227
0.105	4.6	2.2	0.055	4.5	2.1	0.057	4.7	1.9	0.024	4.7	2.0	0.093	9.1	3.7	0.046	9.4	3.4	0.078	18.5	6.5	210
0.060		2.0	0.091	4.8	1.9	0.084	4.9	1.5	0.068	4.9	2.0	0.086	9.7	3.3	0.091	9.8	2.9	0.101	19.6	5.5	209
0.167	3.9	2.1	0.255	4.0	2.2	0.156	4.3	1.7	0.211	4.4	2.1	0.254	7.9	3.5	0.238	8.7	2.9	0.275	16.7	5.8	201
0.076	4.2	2.1	0.086	4.4	2.0	0.088	4.4	1.9	0.014	4.0	2.1	0.095	8.6	3.4	0.058	8.4	3.3	0.089	17.0	5.8	196
0.274	5.2	1.8	0.272	5.1	1.8	0.209	5.2	1.7	0.146	5.3	2.0	0.332	10.3	3.0	0.214	10.5	3.0	0.313	20.7	5.2	191
0.330	4.4	2.2	0.239	4.4	2.0	0.192	4.2	1.9	0.242	4.4	2.0	0.333	8.9	3.6	0.270	8.6	3.1	0.343	17.5	5.9	187
0.024	5.0	1.8	0.076	5.1	1.8	0.173	5.2	1.8	0.078	5.1	1.7	0.059	10.1	3.0	0.161	10.3	2.8	0.124	20.4	5.1	182
0.286	4.3	2.2	0.398	4.7	2.1	0.258	4.7	1.8	0.203	4.7	2.0	0.393	9.0	3.7	0.278	9.4	3.2	0.381	18.4	6.2	178
0.114	4.3	2.3	0.138	4.3	2.1	0.129	4.6	2.0	0.200	4.4	2.0	0.145	8.6	3.8	0.194	9.0	3.4	0.186	17.5	6.6	175
0.238	6.7	1.2	0.181	6.4	1.4	0.127	6.2	1.5	0.236	6.8	1.3	0.258	13.2	2.1	0.216	13.1	2.3	0.272	26.2	3.9	172
0.188	5.2	1.9	0.212	5.4	2.1	0.310	5.1	1.8	0.211	5.2	2.1	0.235	10.6	3.4	0.297	10.4	3.4	0.290	21.0	6.2	166
0.334	4.1	2.3	0.291	4.1	2.3	0.180	4.6	2.0	0.154	4.6	2.1	0.358	8.2	4.0	0.193	9.1	3.6	0.309	17.3	6.9	163
0.245	5.6	1.6	0.227	5.7	1.7	0.187	5.2	1.7	0.250	5.6	1.9	0.279	11.4	2.8	0.273	10.8	2.9	0.318	22.2	4.9	159
0.213	5.6	1.9	0.379	5.3	1.8	0.107	5.4	1.8	0.175	5.4	1.8	0.347	10.9	3.1	0.167	10.7	3.0	0.298	21.6	5.3	154
0.250	4.9	2.3	0.274	4.9	2.2	0.332	4.9	1.8	0.152	5.3	2.0	0.300	9.8	3.9	0.283	10.2	3.2	0.318	20.0	6.5	145
0.345	5.5	1.8	0.153	5.4	1.9	0.257	5.3	1.7	0.307	5.6	1.8	0.291	11.0	3.1	0.344	10.8	2.9	0.359	21.8	5.3	136
0.166	4.2	2.2	0.234	3.8	2.2	0.090	4.4	2.0	0.225	4.2	2.0	0.235	8.0	3.8	0.185	8.6	3.4	0.237	16.6	6.4	132
0.171	5.7	1.7	0.218	5.8	1.6	0.162	5.4	1.7	0.137	5.7	1.6	0.248	11.6	2.6	0.186	11.1	2.6	0.263	22.7	4.3	131
0.171	5.3	2.1	0.329	5.6	1.8	0.184	5.2	1.7	0.132	5.3	2.0	0.290	10.9	3.3	0.187	10.5	3.1	0.271	21.4	5.6	128
0.246	5.1	1.9	0.230	5.2	1.7	0.220	5.3	1.9	0.256	5.4	2.0	0.276	10.4	3.2	0.296	10.7	3.1	0.325	21.0	5.5	126
0.178	5.3	1.8	0.128	4.9	1.9	0.127	4.9	1.9	0.107	5.2	1.9	0.178	10.1	3.2	0.136	10.1	3.3	0.177	20.2	5.7	118
0.161	6.7	1.4	0.318	6.8	1.2	0.226	6.5	1.5	0.205	6.8	1.3	0.282	13.5	2.1	0.256	13.4	2.3	0.302	26.9	4.0	114
0.297	4.8	2.1	0.268	4.7	2.1	0.250	4.9	1.9	0.223	5.0	2.0	0.331	9.5	3.5	0.286	9.8	3.2	0.350	19.3	5.9	110
0.176	1.9	1.9	0.329	1.5	2.0	0.281	2.9	1.9	0.115	2.7	2.0	0.307	3.4	3.3	0.250	5.6	3.1	0.318	9.0	5.6	109
0.409	5.0	2.1	0.377	4.7	2.2	0.285	4.8	2.0	0.318	4.7	2.2	0.450	9.7	3.7	0.368	9.5	3.4	0.484	19.1	6.1	107
0.162	4.2	2.2	0.284	4.7	2.2	0.153	4.4	1.9	0.151	4.4	2.1	0.262	8.9	3.7	0.176	8.8	3.4	0.252	17.7	6.2	104
0.198	5.3	1.7	0.209	5.3	1.7	0.420	5.2	1.6	0.225	5.2	1.6	0.266	10.6	2.6	0.386	10.4	2.7	0.380	21.1	4.5	104
0.175	5.3	2.0	0.448	4.8	1.9	0.245	4.8	2.0	0.318	5.0	2.1	0.368	10.1	3.3	0.326	9.8	3.6	0.388	19.8	6.1	101
0.170	5.3	2.2	0.144	5.2	1.8	0.051	5.1	1.8	0.169	5.3	1.9	0.191	10.5	3.3	0.136	10.4	3.1	0.183	20.9	5.7	100
0.084	4.7	2.1	0.069	4.9	1.6	0.278	4.6	1.9	0.054	5.1	2.1	0.092	9.6	3.1	0.191	9.7	3.4	0.164	19.3	5.6	100
0.142	6.4	1.3	0.057	6.6	1.4	0.136	6.1	1.5	−0.048	6.7	1.2	0.117	12.9	2.2	0.067	12.8	2.2	0.111	25.7	3.7	96
0.371	4.6	2.1	0.312	4.9	1.9	0.292	4.2	2.0	0.347	5.1	2.0	0.405	9.4	3.4	0.373	9.3	3.4	0.439	18.7	6.0	96
0.112	3.8	1.9	0.354	4.5	1.8	−0.002	4.2	1.9	0.325	4.4	2.1	0.288	8.3	3.0	0.202	8.6	3.4	0.288	16.9	5.3	95
0.179	—	—	0.222	—	—	0.184	—	—	0.201	—	—	0.260	—	—	0.221	—	—	0.272	—	—	—

BEST COPY AVAILABLE

Table 4. Correlations of Scores Based on All Items with First-Year Average Grades in Each College

	Formula Scores												*Scaled Scores*									
	Antonyms			*Analogies*			*Reading Comprehension*			*Sentence Completion*			*Vocabulary*			*Reading*			*SAT-V*			*Number of*
College	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*R*	*Mean*	*S.D.*	*Cases*
1	0.257	10.4	4.3	0.290	9.0	3.6	0.293	9.5	4.7	0.260	7.9	2.9	0.296	44	9	0.310	44	9	0.328	441	88	913
2	0.160	11.7	4.3	0.230	10.0	3.2	0.229	10.9	4.8	0.195	8.5	2.8	0.208	47	8	0.241	46	9	0.241	470	88	602
3	0.211	10.3	5.5	0.255	9.0	4.3	0.279	10.0	5.5	0.267	8.2	3.4	0.247	44	11	0.303	45	11	0.287	447	112	592
4	0.185	10.3	4.5	0.219	8.7	3.7	0.181	9.5	5.0	0.148	7.7	3.1	0.211	44	9	0.185	43	10	0.215	438	95	573
5	0.292	8.9	4.3	0.283	7.0	3.6	0.306	7.5	4.7	0.285	6.8	2.8	0.322	40	9	0.338	40	9	0.359	397	84	538
6	0.177	10.5	4.4	0.197	9.0	3.6	0.230	9.6	4.7	0.126	7.9	2.9	0.208	44	9	0.209	44	9	0.225	444	89	469
7	0.327	11.5	4.4	0.2[illegible]	9.0	3.5	0.341	9.9	5.0	0.293	8.5	3.0	0.353	46	9	0.361	[illegible]	10	0.377	456	93	463
8	0.195	10.5	4.3	[illegible]	8.8	3.5	0.235	10.2	4.7	0.186	8.0	2.8	0.217	44	9	0.241	45	9	0.249	446	86	453
9	0.199	9.1	4.4	[illegible]	7.5	3.6	0.215	8.7	4.5	0.153	7.3	2.9	0.222	41	9	0.218	42	9	0.239	413	89	389
10	0.242	9.8	4.6	0.227	8.3	3.6	0.232	8.9	4.8	0.235	7.6	3.2	0.258	43	9	0.266	42	9	0.278	426	93	356
11	0.226	12.3	5.4	0.265	10.6	4.1	0.195	12.4	5.5	0.247	9.3	3.2	0.255	48	11	0.231	50	11	0.258	495	112	278
12	0.181	11.3	4.6	0.250	9.7	3.8	0.294	10.4	5.4	0.190	8.3	3.2	0.231	46	9	0.280	46	11	0.272	461	100	278
13	0.300	10.2	4.6	0.275	8.8	3.6	0.304	9.5	4.9	0.272	7.7	3.0	0.342	44	9	0.320	43	10	0.355	437	92	259
14	0.227	9.8	4.4	0.211	8.1	3.5	0.234	9.0	4.7	0.315	7.5	3.0	0.242	42	9	0.297	43	9	0.293	425	88	235
15	0.073	9.8	3.5	0.022	8.5	2.6	0.044	9.1	3.8	0.011	7.5	2.4	0.058	43	7	0.035	43	7	0.056	429	65	227
16	0.078	10.1	4.6	0.060	8.4	3.7	0.070	8.7	4.9	0.027	7.5	2.8	0.077	43	9	0.061	42	9	0.073	428	93	210
17	−0.015	10.3	4.2	0.009	8.7	3.2	0.133	9.4	4.1	0.037	7.7	2.7	−0.008	44	8	0.104	43	8	0.053	437	78	209
18	0.122	8.8	4.1	0.244	7.4	3.6	0.127	7.8	4.0	0.227	7.0	3.0	0.200	40	9	0.195	40	8	0.215	402	81	201
19	0.071	8.8	3.9	0.067	7.8	3.2	0.146	7.8	4.5	0.009	6.6	2.8	0.082	41	8	0.104	40	8	0.102	402	80	196
20	0.264	11.2	3.9	0.315	9.4	3.3	0.223	10.4	4.6	0.192	8.5	2.7	0.333	46	8	0.238	46	9	0.304	459	80	191
21	0.265	9.3	4.3	0.285	8.1	3.3	0.232	7.6	4.2	0.292	7.2	2.9	0.304	42	8	0.288	40	8	0.321	409	83	187
22	0.080	10.9	3.8	0.113	9.3	3.3	0.148	9.9	4.4	0.094	8.1	2.4	0.111	45	8	0.147	44	8	0.140	450	74	182
23	0.345	9.2	4.4	0.380	8.3	3.6	0.371	8.5	4.5	0.210	7.3	2.9	0.397	42	9	0.358	42	9	0.406	418	87	178
24	0.118	9.1	4.6	0.154	7.9	3.7	0.137	8.4	4.9	0.152	7.1	2.8	0.151	41	9	0.161	41	9	0.170	412	90	175
25	0.328	15.5	3.7	0.359	13.1	3.4	0.333	15.5	4.6	0.210	11.3	2.3	0.377	56	8	0.322	57	9	0.383	570	84	172
26	0.235	11.0	4.3	0.174	9.7	3.5	0.270	10.5	4.7	0.230	8.2	3.0	0.227	46	9	0.290	46	9	0.273	460	89	166
27	0.344	9.2	4.7	0.329	7.7	3.7	0.344	8.3	4.5	0.240	7.3	3.1	0.379	41	9	0.338	41	9	0.385	412	90	163
28	0.284	12.0	3.7	0.349	10.2	3.4	0.266	11.3	4.9	0.330	8.9	2.9	0.343	48	8	0.327	47	9	0.375	480	82	159
29	0.241	11.2	3.9	0.350	9.3	3.3	0.185	10.9	4.4	0.239	8.5	2.5	0.313	46	8	0.227	46	8	0.296	463	81	154
30	0.285	10.4	4.7	0.250	9.2	3.9	0.305	9.9	4.8	0.240	8.4	2.9	0.311	44	9	0.311	45	9	0.334	449	93	145
31	0.303	12.0	4.1	0.186	10.1	3.5	0.348	11.6	4.8	0.333	8.8	2.7	0.269	47	8	0.367	48	9	0.349	480	85	136
32	0.242	9.1	4.2	0.262	7.2	3.5	0.225	7.6	4.7	0.247	7.0	2.7	0.279	40	8	0.259	40	9	0.278	402	87	132
33	0.271	12.0	3.6	0.308	10.5	2.9	0.297	11.5	4.6	0.251	8.9	2.4	0.328	48	7	0.304	48	9	0.353	482	76	131
34	0.210	11.6	4.4	0.357	9.9	3.2	0.270	10.7	4.7	0.240	8.6	2.7	0.299	47	8	0.282	46	9	0.312	468	87	128
35	0.328	11.4	4.4	0.282	9.6	3.6	0.309	10.6	5.3	0.250	8.5	3.0	0.339	46	9	0.324	46	10	0.345	465	97	126
36	0.202	11.1	4.3	0.154	9.4	3.7	0.157	9.4	4.8	0.079	8.4	2.7	0.203	46	9	0.142	44	9	0.191	452	86	118
37	0.291	15.7	3.8	0.363	14.0	3.4	0.296	16.3	4.6	0.279	11.4	2.5	0.347	57	8	0.328	58	9	0.368	584	85	114
38	0.373	10.5	4.2	0.308	8.7	3.7	0.338	9.4	4.8	0.237	8.0	2.8	0.377	44	9	0.344	44	9	0.382	439	89	110
39	0.215	3.8	3.8	0.350	3.5	3.5	0.346	4.0	3.5	0.197	4.5	2.9	0.326	30	7	0.333	32	7	0.361	301	71	109
40	0.404	10.7	4.3	0.423	8.8	3.8	0.474	9.4	4.5	0.354	7.9	2.9	0.457	44	9	0.469	44	9	0.517	442	86	107
41	0.186	9.2	4.5	0.260	8.5	3.4	0.197	7.8	3.9	0.146	7.0	3.0	0.239	42	9	0.197	40	8	0.245	412	83	104
42	0.330	11.1	3.8	0.192	9.3	3.3	0.429	11.0	4.5	0.274	8.2	2.5	0.317	45	8	0.421	46	8	0.414	460	76	104
43	0.190	11.4	4.2	0.479	9.1	3.5	0.343	9.9	5.2	0.345	8.1	3.1	0.372	46	8	0.385	45	10	0.405	453	91	101
44	0.122	11.2	4.2	0.082	9.6	3.8	0.134	10.3	4.6	0.169	8.3	2.7	0.118	46	9	0.154	45	9	0.155	459	88	100
45	0.086	10.2	4.1	0.108	8.7	3.0	0.162	9.2	4.6	0.089	8.0	3.0	0.102	44	8	0.151	44	9	0.143	437	82	100
46	0.080	14.5	3.4	0.051	12.7	3.1	0.119	14.2	3.7	0.075	10.8	2.1	0.074	54	7	0.114	54	7	0.096	547	67	96
47	0.422	10.3	4.0	0.443	8.6	3.3	0.351	8.3	4.4	0.326	8.1	3.1	0.493	43	8	0.383	42	9	0.472	430	80	96
48	0.121	8.7	3.8	0.325	8.2	3.0	0.159	7.4	4.2	0.313	6.9	2.7	0.250	41	7	0.250	40	8	0.267	403	72	95
Median	0.226	—	—	0.257	—	—	0.235	—	—	0.236	—	—	0.263	—	—	0.281	—	—	0.290	—	—	—

BEST COPY AVAILABLE

Results shown in Table 4 can be summarized by counting the number of colleges in which each item type has the highest validity and the number in which each item type has the lowest validity. Results are as follows:

	Lowest	Highest
Antonyms (25 items)	12	6½
Analogies (20 items)	8	21
Reading Comprehension (25 items)	5	17½
Sentence Completion (15 items)	23	3
Total Colleges	48	48

The higher validity for analogies and reading comprehension is consistent with the results shown in Table 2 based on all test items.

When the vocabulary and reading subscores are compared for each college, using results from Table 4, vocabulary has a higher validity in 23 of the 48 comparisons, and there are two ties. The median validity coefficient is higher for reading (.281) than for vocabulary (.263).

On the whole, the variations in relative validity of the four item types in the different analyses suggest that there may not be much difference in validity between one item type and another. It is true that when matched subsets are compared, analogies and antonyms tend to have somewhat higher validities than reading comprehension and sentence completion. On the other hand, when results for all items are considered, reading comprehension and analogies tend to have somewhat higher validities than antonyms and sentence completion. The difference in results between the matched subsets and the whole set of items of each type may have arisen from a number of factors: (1) all items in the matched subsets were attempted by a large percentage of students, (2) the subsets were matched with respect to level and dispersion of item difficulties, (3) the particular items selected for the subset happened to have relatively high or low validity as compared with the items not selected, or (4) some combination of these and other factors.

The small differences in validity between vocabulary and reading subscores and the fact that vocabulary scores based on matched subsets of items tend to have higher validity than the corresponding reading subscores are not consistent with the marked superiority for reading over vocabulary found by Ramist. Because Ramist's findings were based on intact student groups and on a number of forms of the SAT-verbal sections, the difference in results may be attributable to the characteristics of the test form used in this study, to the fact that only December 1980 examinees were included in the sample for this study and to other differences between the samples, and to sampling fluctuations.

Table 5 shows intercorrelation and validities based on pooled data across all 48 colleges, giving equal weight to each of the 11,320 students included in the sample. Within the matched subsets, the highest correlation (.502) is between antonyms and sentence completion and the lowest (.363) is between analogies and reading comprehension. For the intercorrelations of part scores based on all items, the highest correlation (.625) is between antonyms and analogies. The correlations shown in Table 5 between matched subset scores and corresponding scores based on all items are inflated to some extent because of the overlap between the matched subsets and the set of all items of each item type. A similar effect increases the correlations relating any one of the item types to the vocabulary and reading subscores and all correlations relating SAT-verbal scores to other verbal scores.

Tables 6 and 7 show validity coefficients and intercorrelations of the various SAT-verbal subscores for four subgroups of students. Results for the matched subsets are of particular interest. For the four subgroups (men, women, students in higher scoring colleges, and students in lower scoring colleges), the highest correlation is between sentence completion and antonyms. The lowest correlation is between reading comprehension and analogies except for students in the higher scoring colleges, for which reading comprehension and antonyms have a slightly lower coefficient. These results are consistent with those of the survey of item and test analysis results included as the first study in this report. The survey showed that, on the basis of true-score intercorrelations for 20 recent forms of the SAT-verbal sections, sentence completion and antonyms had the highest median correlation and reading comprehension and analogies had the lowest median correlation.

The relatively small differences in validity for the four item types and the marked variation in the relative validities from one college to another as shown in Tables 2, 3, and 4 limit the value of the multiple correlation analysis. Nevertheless, it is useful to know the extent to which optimal weighting of each item type within each college increases predictive validity when scores on the four item types are used in combination with each other and when they are used in combination with other predictors. In this study, SAT-verbal, SAT-mathematical, and Test of Standard Written English (TSWE) scores, and the measure of high school record (HSR) used by the college in its validity study were used in the multiple correlation analyses. The validity, mean, and standard deviation of these predictors are shown in Table 8.

The median validity of total score on the four matched subsets is .272 as shown in Table 3. It is useful to compare this figure with the multiple correlations for combinations of these scores shown in Table 9. When reading and vocabulary subscores based on matched item subsets are optimally weighted, the median multiple correlation is .276. When each item type part score is optimally weighted, the median multiple correlation is .295. The small size of this

Table 5. Intercorrelations of Scores Based on Matched Item-Type Subsets, Scores Based on All Items, SAT-Verbal, SAT-Mathematical, High School Record, and First-Year Average Grades (Based on Pooled Standardized Covariances for 48 Colleges)

	Scores Based on Matched Subsets							*Scores Based on All Items*							*Other Predictors*			
	Antonyms	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Vocabulary*	*Reading*	*Total*	*Antonyms*	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Vocabulary*	*Reading*	*SAT-V*	*SAT-M*	*TSWE*	*HSR*	*First-Year Average*
Matched Subsets																		
Antonyms	1.000	0.471	0.397	0.502	0.864	0.538	0.792	0.835	0.540	0.528	0.546	0.775	0.595	0.740	0.380	0.557	0.191	0.202
Analogies	0.471	1.000	0.363	0.426	0.850	0.472	0.748	0.516	0.826	0.466	0.463	0.723	0.518	0.670	0.425	0.531	0.204	0.212
Reading Comprehension	0.397	0.363	1.000	0.403	0.443	0.823	0.702	0.444	0.430	0.750	0.440	0.483	0.704	0.636	0.336	0.443	0.191	0.186
Sentence Completion	0.502	0.426	0.403	1.000	0.542	0.851	0.775	0.553	0.506	0.520	0.898	0.588	0.737	0.712	0.387	0.526	0.165	0.187
Vocabulary	0.864	0.850	0.443	0.542	1.000	0.589	0.899	0.791	0.792	0.580	0.589	0.874	0.649	0.822	0.468	0.634	0.231	0.241
Reading	0.538	0.472	0.823	0.851	0.589	1.000	0.883	0.596	0.560	0.752	0.808	0.640	0.861	0.806	0.433	0.580	0.212	0.222
Total	0.792	0.748	0.702	0.775	0.899	0.883	1.000	0.781	0.762	0.743	0.779	0.853	0.843	0.914	0.506	0.681	0.248	0.260
All Items																		
Antonyms	0.835	0.516	0.444	0.553	0.791	0.596	0.781	1.000	0.625	0.593	0.616	0.920	0.671	0.860	0.401	0.623	0.205	0.219
Analogies	0.540	0.826	0.430	0.506	0.792	0.560	0.762	0.625	1.000	0.563	0.568	0.877	0.630	0.815	0.439	0.598	0.234	0.243
Reading Comprehension	0.528	0.466	0.750	0.520	0.580	0.752	0.743	0.593	0.563	1.000	0.579	0.641	0.937	0.847	0.454	0.569	0.250	0.246
Sentence Completion	0.546	0.463	0.440	0.898	0.589	0.808	0.779	0.616	0.568	0.579	1.000	0.657	0.823	0.795	0.419	0.575	0.191	0.211
Vocabulary	0.775	0.723	0.483	0.588	0.874	0.640	0.853	0.920	0.877	0.641	0.657	1.000	0.722	0.930	0.462	0.675	0.242	0.254
Reading	0.595	0.518	0.704	0.737	0.649	0.861	0.843	0.671	0.630	0.937	0.823	0.722	1.000	0.923	0.492	0.636	0.254	0.260
SAT-V	0.740	0.670	0.636	0.712	0.822	0.806	0.914	0.860	0.815	0.847	0.795	0.930	0.923	1.000	0.514	0.706	0.266	0.277
SAT-M	0.380	0.425	0.336	0.387	0.468	0.433	0.506	0.401	0.439	0.454	0.419	0.462	0.492	0.514	1.000	0.461	0.322	0.294
TSWE	0.577	0.531	0.443	0.526	0.634	0.580	0.681	0.623	0.598	0.569	0.575	0.675	0.636	0.706	0.461	1.000	0.288	0.285
HSR	0.191	0.204	0.191	0.165	0.231	0.212	0.248	0.205	0.234	0.250	0.191	0.242	0.254	0.266	0.322	0.288	1.000	0.415
First-Year Average	0.202	0.212	0.186	0.187	0.241	0.222	0.260	0.219	0.243	0.246	0.211	0.254	0.260	0.277	0.294	0.285	0.415	1.000

Table 6. Intercorrelations of Scores Based on Matched Item-Type Subsets, Scores Based on All Items, SAT-Verbal, SAT-Mathematical, High School Record, and First-Year Average Grades, for Men and Women Separately* (Mean Equally-Weighted Standardized Covariances)

	Scores Based on Matched Subsets							*Scores Based on All Items*						*Other Predictors*				
	Antonyms	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Vocabulary*	*Reading*	*Total*	*Antonyms*	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Vocabulary*	*Reading*	*SAT-V*	*SAT-M*	*TSWE*	*HSR*	*First-Year Average*
Matched Subsets																		
Antonyms		.466	.394	.502	.864	.538	.794	.835	.545	.526	.551	.779	.598	.742	.389	.556	.186	.219
Analogies	.466		.353	.413	.846	.461	.740	.508	.821	.465	.461	.718	.517	.666	.424	.524	.219	.210
Reading Comprehension	.410	.360		.387	.436	.818	.697	.445	.421	.746	.426	.481	.696	.630	.350	.437	.210	.176
Sentence Completion	.485	.423	.398		.535	.845	.769	.547	.501	.511	.897	.583	.734	.706	376	.540	.168	.203
Vocabulary	.857	.854	.448	.530		.583	.897	.790	.792	.580	.592	.875	.652	.823	.473	.630	.236	.249
Reading	.536	.469	.822	.848	.585		.882	.597	.555	.750	.804	.640	.859	.804	.436	.589	.225	.228
Total	.789	.751	.703	.765	.899	.880		.782	.761	.743	.779	.856	.845	.915	.511	.684	.258	.267
All Items																		
Antonyms	.835	.505	.455	.544	.783	.598	.779		.620	.600	.618	.918	.678	.861	.415	.625	.202	.234
Analogies	.540	.823	.418	.499	.795	.549	.761	.623		.562	.566	.877	.630	.813	.438	.592	.231	.240
Reading Comprehension	.527	.456	.754	.515	.573	.753	.739	.595	.554		.573	.646	.935	.847	.461	.572	.278	.254
Sentence Completion	.539	.460	.442	.889	.583	.805	.773	.614	.572	.577		.659	.822	.794	.411	.591	.193	.227
Vocabulary	.778	.714	.483	.579	.872	.636	.852	.921	.874	.636	.659		.728	.931	.470	.676	.237	.262
Reading	.592	.508	.711	.725	.641	.859	.836	.671	.624	.940	.816	.718		.924	.495	.646	.275	.272
SAT-V	.739	.661	.642	.701	.817	.804	.911	.859	.810	.848	.793	.927	.923		.519	.710	.274	.286
SAT-M	.374	.431	.314	.375	.470	.413	.497	.390	.443	.433	.403	.456	.469	.499		.529	.386	.329
TSWE	.553	.528	.452	.523	.631	.584	.683	.612	.593	.578	.575	.665	.640	.705	.472		.281	.289
HSR	.190	.223	.163	.159	.242	.190	.243	.198	.248	.226	.190	.245	.237	.259	.369	.255		.422
First-Year Average	.202	.223	.177	.186	.249	.217	.262	.219	.254	.238	.216	.258	.256	.278	.319	.254	.408	

*Results for women are above the diagonal; results for men are below the diagonal. Only colleges having at least 50 women and 50 men are included.

Table 7. Intercorrelations of Scores Based on Matched Item-Type Subsets, Scores Based on All Items, SAT-Verbal, SAT-Mathematical, High School Record, and First-Year Average Grades for Seven Colleges with Relatively High SAT-V Means and Seven Colleges with Relatively Low SAT-V Means* (Based on Pooled Standardized Covariances)

	Scores Based on Matched Subsets							*Scores Based on All Items*							*Other Predictors*			
	Antonyms	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Vocabulary*	*Reading*	*Total*	*Antonyms*	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Vocabulary*	*Reading*	*SAT-V*	*SAT-M*	*TSWE*	*HSR*	*First-Year Average*
Matched Subsets																		
Antonyms		.425	.346	.472	.844	.490	.754	.757	.533	.482	.537	.711	.553	.682	.277	.509	.176	.227
Analogies	.431		.352	.401	.841	.453	.732	.456	.768	.436	.450	.653	.485	.617	.314	.472	.191	.213
Reading Comprehension	.337	.285		.374	.409	.831	.703	.418	.404	.724	.433	.444	.685	.610	.290	.445	.213	.176
Sentence Completion	.465	.415	.342		.512	.825	.757	.538	.526	.498	.857	.576	.688	.684	.323	.523	.106	.204
Vocabulary	.846	.845	.367	.520		.553	.880	.715	.769	.540	.582	.806	.612	.767	.350	.577	.218	.258
Reading	.492	.432	.805	.832	.546		.881	.573	.559	.737	.776	.612	.828	.779	.372	.583	.192	.230
Total	.770	.735	.656	.761	.890	.867		.728	.782	.723	.769	.804	.816	.877	.410	.656	.232	.277
All Items																		
Antonyms	.820	.490	.392	.531	.775	.566	.768		.659	.576	.629	.922	.656	.856	.296	.575	.163	.260
Analogies	.489	.849	.359	.480	.790	.517	.749	.570		.546	.618	.894	.632	.829	.338	.586	.215	.279
Reading Comprehension	.470	.426	.748	.466	.530	.735	.714	.554	.511		.587	.611	.945	.841	.410	.551	.217	.263
Sentence Completion	.500	.459	.365	.905	.567	.786	.763	.578	.538	.515		.679	.812	.806	.341	.569	.145	.253
Vocabulary	.754	.735	.426	.574	.880	.614	.856	.906	.860	.908	.633		.702	.923	.341	.630	.203	.292
Reading	.548	.502	.680	.727	.621	.860	.836	.641	.596	.940	.602	.701		.918	.426	.613	.211	.285
SAT-V	.710	.674	.594	.703	.818	.795	.918	.844	.793	.820	.774	.926	.915		.415	.672	.223	.315
SAT-M	.380	.459	.326	.405	.495	.448	.537	.395	.472	.447	.442	.483	.505	.535		.335	.301	.351
TSWE	.502	.512	.393	.496	.599	.545	.652	.584	.575	.530	.548	.653	.612	.687	.495		.225	.275
HSR	.171	.228	.154	.153	.236	.187	.241	.189	.222	.207	.157	.298	.212	.241	.333	.282		.417
First-Year Average	.205	.238	.157	.209	.263	.225	.278	.214	.255	.236	.232	.262	.267	.287	.303	.294	.436	

*Results for colleges with relatively high SAT-V means are above the diagonal; results for colleges with relatively low SAT-V means are below the diagonal.

Table 8. Correlations of SAT Scores and High School Record with First-Year Average Grades in Each College

College	SAT-Verbal			SAT-Mathematical			TSWE			High School Record			Number of Cases
	R	Mean	S.D.	R	Mean	S.D.	R	Mean	S.D.	R	Mean	S.D.	
1	0.328	441	88	0.356	495	100	0.358	44	9	0.369	57.8	6.3	913
2	0.241	470	88	0.277	524	90	0.262	47	8	0.473	3.0	0.5	602
3	0.287	447	112	0.375	527	108	0.281	45	10	0.484	3.4	0.4	592
4	0.215	438	95	0.281	493	104	0.240	44	10	0.382	3.1	0.5	573
5	0.359	397	84	0.317	427	100	0.388	40	10	0.497	55.5	7.8	538
6	0.225	444	89	0.273	500	95	0.222	45	9	0.342	56.3	6.4	469
7	0.377	456	93	0.407	490	99	0.382	46	9	0.523	3.3	0.4	463
8	0.249	446	86	0.269	526	99	0.300	45	9	0.359	59.6	5.7	453
9	0.239	413	89	0.146	439	93	0.227	43	10	0.266	57.5	7.7	389
10	0.278	426	93	0.265	478	95	0.297	43	10	0.473	3.2	0.4	356
11	0.258	495	112	0.307	587	103	0.237	49	10	0.369	3.6	0.4	278
12	0.272	461	100	0.259	515	98	0.202	46	9	0.432	3.2	0.5	278
13	0.355	437	92	0.338	498	100	0.355	44	9	0.605	3.3	0.4	259
14	0.293	425	88	0.252	470	91	0.328	44	9	0.387	3.1	0.5	235
15	0.056	429	65	0.199	478	72	0.136	44	8	0.255	3.1	0.4	227
16	0.073	428	93	0.084	477	95	0.096	44	10	0.386	3.2	0.4	210
17	0.053	437	78	0.098	501	83	0.038	45	9	0.457	3.0	0.5	209
18	0.215	402	81	0.209	435	87	0.184	41	9	0.380	2.8	0.4	201
19	0.102	402	80	0.336	434	73	0.139	43	8	0.358	3.0	0.5	196
20	0.304	459	80	0.229	508	94	0.308	47	8	0.378	58.5	6.5	191
21	0.321	409	83	0.211	449	83	0.316	41	10	0.299	2.8	0.5	187
22	0.140	450	74	0.161	522	90	0.210	46	8	0.292	3.1	0.5	182
23	0.406	418	87	0.335	451	88	0.412	43	9	0.405	2.9	0.5	178
24	0.170	412	90	0.253	480	110	0.190	42	10	0.294	3.1	0.5	175
25	0.383	570	84	0.409	629	84	0.305	55	6	0.354	3.7	0.3	172
26	0.273	460	89	0.236	541	93	0.271	45	9	0.279	59.6	6.5	166
27	0.385	412	90	0.349	463	106	0.295	42	9	0.545	3.1	0.4	163
28	0.375	480	82	0.362	537	80	0.296	48	8	0.511	3.3	0.5	159
29	0.296	463	81	0.359	530	86	0.327	48	8	0.440	60.8	6.5	154
30	0.334	449	93	0.336	513	104	0.361	45	9	0.466	58.7	5.9	145
31	0.349	480	85	0.540	537	96	0.345	48	8	0.591	60.6	8.5	136
32	0.278	402	87	0.306	429	96	0.263	41	9	0.607	3.0	0.5	132
33	0.353	482	76	0.210	541	78	0.297	48	8	0.345	59.3	6.2	131
34	0.312	468	87	0.333	516	95	0.419	48	8	0.422	3.2	0.5	128
35	0.345	465	97	0.378	485	91	0.349	46	10	0.589	3.1	0.7	126
36	0.191	452	86	0.132	492	90	0.216	44	9	0.363	53.9	5.3	118
37	0.368	584	85	0.325	646	70	0.250	55	6	0.353	66.9	7.6	114
38	0.382	439	89	0.334	459	74	0.356	46	8	0.450	58.7	7.2	110
39	0.361	301	71	0.419	338	71	0.390	31	9	0.486	55.9	7.9	109
40	0.517	442	86	0.392	468	97	0.336	46	9	0.470	57.6	7.9	107
41	0.245	412	83	0.148	458	88	0.293	43	9	0.372	2.9	0.5	104
42	0.414	460	76	0.298	514	83	0.396	47	7	0.431	3.3	0.5	104
43	0.405	453	91	0.487	516	104	0.492	46	10	0.555	3.2	0.5	101
44	0.155	459	88	0.248	504	105	0.232	46	10	0.353	58.4	6.7	100
45	0.143	437	87	0.202	482	84	0.258	45	8	0.364	57.3	5.6	100
46	0.096	547	67	0.314	583	70	0.201	54	6	0.440	3.5	0.5	96
47	0.472	430	80	0.436	452	88	0.510	43	9	0.481	52.9	7.9	96
48	0.267	403	72	0.391	444	80	0.209	41	9	0.345	52.9	5.7	95
Median	0.290	—	—	0.307	—	—	0.295	—	—	0.396	—	—	—

Table 9. Multiple Correlations with First-Year Average Grades in Each College Using Item-Type Scores Based on Matched Subsets of Items

College	Combination of Predictors: Antonyms, Analogies, Reading Compr., Sent. Completion	Antonyms, Analogies, Reading Compr., Sent. Completion, SAT-M	Antonyms, Analogies, Reading Compr., Sent. Completion, SAT-M, HSR	Antonyms, Analogies, Reading Compr., Sent. Completion, TSWE	Vocabulary, Reading	Vocabulary, Reading, SAT-M	Vocabulary, Reading, SAT-M, HSR	Vocabulary, Reading, TSWE	Total* SAT-M, HSR	Number of Cases
1	0.340	0.392	0.455	0.380	0.336	0.390	0.453	0.377	0.453	913
2	0.220	0.296	0.489	0.277	0.213	0.291	0.486	0.270	0.485	602
3	0.269	0.396	0.542	0.296	0.265	0.389	0.539	0.292	0.537	592
4	0.211	0.289	0.425	0.251	0.205	0.286	0.422	0.245	0.422	573
5	0.328	0.361	0.520	0.397	0.325	0.359	0.516	0.395	0.516	538
6	0.258	0.309	0.403	0.272	0.241	0.294	0.392	0.253	0.392	469
7	0.366	0.445	0.593	0.409	0.363	0.442	0.590	0.407	0.586	463
8	0.239	0.301	0.398	0.318	0.238	0.297	0.394	0.313	0.390	453
9	0.205	0.207	0.312	0.240	0.197	0.199	0.307	0.233	0.306	389
10	0.259	0.313	0.508	0.305	0.259	0.312	0.507	0.305	0.505	356
11	0.282	0.359	0.459	0.283	0.276	0.356	0.454	0.278	0.447	278
12	0.223	0.309	0.464	0.255	0.166	0.266	0.440	0.205	0.439	278
13	0.336	0.388	0.611	0.374	0.333	0.386	0.610	0.372	0.609	259
14	0.283	0.317	0.454	0.342	0.268	0.303	0.438	0.335	0.437	235
15	0.115	0.232	0.315	0.169	0.057	0.200	0.294	0.139	0.294	227
16	0.110	0.122	0.393	0.122	0.095	0.107	0.392	0.108	0.388	210
17	0.112	0.130	0.465	0.116	0.101	0.121	0.463	0.106	0.462	209
18	0.284	0.293	0.450	0.285	0.275	0.284	0.444	0.276	0.444	201
19	0.123	0.348	0.479	0.158	0.096	0.339	0.474	0.141	0.472	196
20	0.349	0.358	0.465	0.389	0.336	0.345	0.454	0.370	0.446	191
21	0.357	0.358	0.421	0.383	0.348	0.349	0.414	0.374	0.413	187
22	0.184	0.220	0.341	0.267	0.163	0.204	0.332	0.246	0.311	182
23	0.427	0.449	0.586	0.464	0.397	0.425	0.575	0.443	0.554	178
24	0.214	0.284	0.368	0.230	0.196	0.266	0.352	0.216	0.348	175
25	0.298	0.456	0.504	0.347	0.276	0.423	0.474	0.329	0.472	172
26	0.321	0.337	0.413	0.332	0.300	0.321	0.406	0.313	0.405	166
27	0.371	0.421	0.595	0.380	0.362	0.407	0.589	0.372	0.572	163
28	0.319	0.398	0.583	0.337	0.319	0.398	0.576	0.337	0.576	159
29	0.384	0.450	0.554	0.421	0.347	0.421	0.532	0.381	0.508	154
30	0.370	0.411	0.517	0.406	0.319	0.376	0.504	0.370	0.501	145
31	0.397	0.566	0.667	0.423	0.365	0.557	0.656	0.395	0.644	136
32	0.275	0.330	0.623	0.306	0.242	0.317	0.619	0.285	0.618	132
33	0.275	0.301	0.413	0.324	0.269	0.299	0.413	0.320	0.413	131
34	0.343	0.413	0.491	0.455	0.291	0.374	0.465	0.431	0.462	128
35	0.325	0.399	0.629	0.362	0.325	0.398	0.627	0.362	0.623	126
36	0.196	0.205	0.418	0.233	0.183	0.188	0.403	0.226	0.401	118
37	0.337	0.395	0.468	0.358	0.304	0.377	0.450	0.320	0.442	114
38	0.358	0.410	0.548	0.385	0.352	0.407	0.541	0.380	0.539	110
39	0.394	0.481	0.594	0.454	0.323	0.444	0.565	0.399	0.562	109
40	0.497	0.528	0.572	0.499	0.488	0.518	0.569	0.490	0.567	107
41	0.292	0.297	0.407	0.334	0.265	0.271	0.399	0.316	0.388	104
42	0.435	0.457	0.538	0.494	0.397	0.412	0.507	0.459	0.500	104
43	0.488	0.542	0.654	0.552	0.391	0.500	0.622	0.498	0.621	101
44	0.205	0.272	0.375	0.247	0.192	0.265	0.370	0.238	0.365	100
45	0.287	0.326	0.450	0.361	0.192	0.236	0.411	0.283	0.399	100
46	0.217	0.353	0.524	0.270	0.120	0.321	0.517	0.205	0.513	96
47	0.441	0.517	0.635	0.520	0.440	0.516	0.634	0.520	0.634	96
48	0.454	0.501	0.541	0.459	0.302	0.414	0.462	0.311	0.461	95
Median *R*	0.295	0.358	0.484	0.340	0.276	0.347	0.464	0.318	0.462	—

*Total based on matched subsets of items.

Table 10. Multiple Correlations with First-Year Average Grades in Each College Using Item-Type Scores Based on All Items

	Combination of Predictors									
College	Antonyms, Analogies, Reading Compr., Sent. Completion	Antonyms, Analogies, Reading Compr., Sent. Completion, SAT-M	Antonyms, Analogies, Reading Compr., Sent. Completion, SAT-M, HSR	Antonyms, Analogies, Reading Compr., Sent. Completion, TSWE	Vocabulary, Reading	Vocabulary, Reading, SAT-M	Vocabulary, Reading, SAT-M, HSR	Vocabulary, Reading, TSWE	SAT-V SAT-M, HSR	Number of Cases
1	0.334	0.388	0.450	0.376	0.327	0.384	0.448	0.373	0.448	913
2	0.266	0.317	0.490	0.300	0.246	0.302	0.486	0.281	0.485	602
3	0.307	0.399	0.544	0.321	0.303	0.399	0.542	0.318	0.540	592
4	0.299	0.295	0.425	0.259	0.214	0.289	0.422	0.247	0.422	573
5	0.361	0.383	0.528	0.408	0.359	0.382	0.526	0.407	0.525	538
6	0.247	0.301	0.397	0.263	0.226	0.286	0.387	0.241	0.387	469
7	0.378	0.448	0.593	0.411	0.379	0.448	0.593	0.410	0.593	463
8	0.255	0.300	0.390	0.312	0.251	0.298	0.389	0.309	0.390	453
9	0.246	0.246	0.333	0.261	0.237	0.237	0.327	0.251	0.328	389
10	0.277	0.322	0.511	0.309	0.279	0.322	0.511	0.310	0.510	356
11	0.276	0.360	0.457	0.279	0.258	0.351	0.450	0.264	0.442	278
12	0.308	0.338	0.479	0.308	0.280	0.315	0.465	0.281	0.457	278
13	0.359	0.405	0.614	0.383	0.355	0.404	0.612	0.380	0.612	259
14	0.321	0.338	0.457	0.357	0.299	0.320	0.446	0.343	0.447	235
15	0.083	0.217	0.307	0.151	0.058	0.204	0.299	0.142	0.295	227
16	0.095	0.111	0.390	0.114	0.077	0.095	0.389	0.097	0.387	210
17	0.170	0.183	0.473	0.170	0.149	0.165	0.471	0.149	0.458	209
18	0.267	0.281	0.445	0.271	0.214	0.240	0.417	0.222	0.417	201
19	0.166	0.360	0.482	0.195	0.105	0.340	0.473	0.142	.473	196
20	0.339	0.350	0.452	0.369	0.333	0.345	0.451	0.357	0.440	191
21	0.331	0.332	0.401	0.359	0.321	0.323	0.393	0.350	0.393	187
22	0.159	0.188	0.316	0.235	0.150	0.180	0.314	0.219	0.311	182
23	0.447	0.461	0.574	0.474	0.411	0.431	0.559	0.444	0.551	178
24	0.178	0.262	0.351	0.204	0.169	0.257	0.346	0.195	0.346	175
25	0.404	0.483	0.519	0.409	0.384	0.477	0.514	0.392	0.510	172
26	0.295	0.321	0.403	0.314	0.291	0.310	0.391	0.312	0.389	166
27	0.415	0.447	0.594	0.415	0.388	0.419	0.586	0.388	0.585	163
28	0.392	0.440	0.601	0.393	0.368	0.426	0.586	0.373	0.588	159
29	0.358	0.430	0.528	0.393	0.313	0.399	0.516	0.355	0.505	154
30	0.334	0.380	0.501	0.379	0.332	0.379	0.501	0.373	0.500	145
31	0.399	0.565	0.668	0.423	0.368	0.555	0.646	0.402	0.644	136
32	0.295	0.342	0.641	0.309	0.288	0.339	0.642	0.305	0.639	132
33	0.357	0.366	0.454	0.366	0.347	0.360	0.450	0.361	0.452	131
34	0.369	0.419	0.497	0.452	0.309	0.378	0.467	0.422	0.468	128
35	0.352	0.416	0.640	0.380	0.352	0.410	0.638	0.377	0.637	126
36	0.227	0.236	0.423	0.258	0.203	0.207	0.416	0.236	0.403	118
37	0.383	0.433	0.494	0.384	0.370	0.421	0.486	0.370	0.481	114
38	0.407	0.445	0.575	0.419	0.390	0.430	0.558	0.407	0.556	110
39	0.400	0.488	0.590	0.451	0.355	0.463	0.575	0.407	0.575	109
40	0.523	0.539	0.587	0.523	0.509	0.528	0.580	0.509	0.585	107
41	0.275	0.281	0.400	0.326	0.243	0.250	0.386	0.305	0.385	104
42	0.447	0.452	0.533	0.488	0.425	0.428	0.510	0.471	0.511	104
43	0.521	0.568	0.668	0.584	0.406	0.504	0.617	0.496	0.618	101
44	0.175	0.262	0.378	0.255	0.154	0.250	0.369	0.242	0.367	100
45	0.164	0.226	0.398	0.282	0.151	0.213	0.395	0.271	0.394	100
46	0.126	0.317	0.513	0.222	0.116	0.315	0.513	0.214	0.513	96
47	0.500	0.548	0.648	0.542	0.504	0.550	0.647	0.540	0.640	96
48	0.436	0.500	0.542	0.440	0.275	0.408	0.464	0.281	0.461	95
Median *R*	0.326	0.360	0.492	0.358	0.301	0.348	0.472	0.330	0.470	—

increment makes it quite unlikely that a shift in the relative weights of the four item types in the SAT-verbal sections would yield a useful increase in validity, in view of the marked variation from one college to another in the relative validity of the item types. Another set of multiple correlations involves the use of SAT-mathematical scores and high school record as additional predictors. In this analysis, the use of these predictors along with the total of the matched scores yields a median multiple correlation of .462. When the total score is replaced by vocabulary and reading subscores based on matched items, the multiple correlation is .464. When the subscores are replaced by the four separate item type scores, the multiple correlation is .484. On the whole, the relatively small differences in the various correlations would not appear to justify shifting the relative weights of the four item types.

The median validity of SAT-verbal scores, based on all 85 verbal items, is .290, as shown in Table 4. Comparison of this figure with multiple correlations based on subscores based on all items of each type shows the effect of optimal weighting on validity. The median multiple correlation based on vocabulary and reading subscores is .301, while the corresponding coefficient based on the four item-type scores is .326, as shown in Table 10. When the SAT-mathematical scores and high school record are included as predictors, the combination including SAT-verbal scores yields a median multiple correlation of .470. Replacing SAT-verbal scores with vocabulary and reading subscores yields a median value of .472, and replacing the two subscores with the four item-type subscores yields a median value of .492. Thus, the results based on all items in the form studied, like those for the matched subsets of items, suggest that modifying the relative weight for the four item types would do little to enhance validity.

DISCUSSION AND CONCLUSIONS

The three studies included in this report were initiated in order to explore the implications of Ramist's findings for the content and format of SAT-verbal sections. The first study surveyed item and test analysis results for the four item types of which the SAT-verbal sections are composed in order to provide information on how those item types are used in meeting test specifications for biserial correlations and difficulty distributions. Results made it clear that, in the forms surveyed, the distribution of item difficulties was different for the four items types. The results for items attempted suggested the possibility that because all four item types were represented in each separately timed section, examinees might differ in allocating their time to different item types.

The second study, using data from the College Board 1981 High School Senior File, was concerned with the relationship of vocabulary and reading subscores to Achievement Test scores, self-reported high school grades in six subject-matter areas, and high school rank. It was done in order to learn more about similarities and differences in the abilities measured by the two subscores. Results indicated that reading had a somewhat higher correlation than vocabulary with nearly all the measures of high school performance included in the study. If the SAT-verbal score is used along with other predictors, however, the results indicated that revising the relative weight given to reading and vocabulary item types would have little effect on the multiple correlations with high school grades.

The present study included two analyses. One was based on subsets of items all of which were answered by 85 percent or more of the validity study sample and which were approximately equal in distribution of item difficulty for four item types, and the other was based on all items.

The results of this study tend to support the view that the four item types have similar validities. Although the limitations of the study, particularly the fact that it was based on only one test form, preclude drawing firm conclusions, the results suggest that changing the amount of time allocated to different types in order to change their relative weights in defining the score is not an especially promising approach to enhancing the validity of the SAT-verbal score. From a methodological viewpoint, the differences between results for the matched subsets and results based on all items of each type suggest that, in possible future studies, the approach of using matched subsets of items as well as all items should be used; and that, if possible, one separately timed section should be devoted to reading comprehension items and the other separately timed section should be devoted to the other three item types. Although the use of separately timed sections would be difficult administratively, it would control the allocation of time by examinees to passage-type items and would thus permit a more rigorous comparison of validities.

In view of the important role of SAT scores in college admissions and of the valuable contribution that an increase in the predictive validity of the SAT would make, further study of the validity of item types seems justified. Future studies might well include analyses of the validity of each item and efforts to relate validity to other characteristics of the item, including difficulty level and biserial correlation with total score. Validity of pretest items might be considered in test assembly. The availability of criterion data through the Validity Study Service and the resources provided by modern computers should make continuing studies of item and item-type validities relatively easy to do. Results of the present study indicate that further studies are needed in order to evaluate the potential usefulness of item-type validity studies for enhancing SAT validity.

REFERENCES

Dressel, P.L. 1940. Some remarks on the Kuder-Richardson reliability coefficient. *Psychometrika* 5, 305–310.

Educational Testing Service. 1955. *Statistical Studies of Selective Service Testing: 1951-1953.* Statistical Report SR-55-30. Princeton, N.J.: Educational Testing Service.

Leary, L.F. and Dorans, N.J. 1982. *The Effects of Item Rearrangement on Test Performance: A Review of the Literature.* Research Report 82-30. Princeton, N.J.: Educational Testing Service.

Olsen, M. and Schrader, W.B. 1953. *A Comparison of Item Types in the College Board Scholastic Aptitude Test.* Statistical Report SR-53. Princeton, N.J.: Educational Testing Service.

———1957. *The Predictive Effectiveness of Four Part Scores Composed of Scholastic Aptitude Test Items.* Statistical Report SR-57-18. Princeton, N.J.: Educational Testing Service.

Pitcher, B. 1960. *The Effectiveness of Tests of Directed Memory, Organization of Ideas and Data Sufficiency for Predicting Business School Grades.* Statistical Report SR-57-3. Princeton, N.J.: Educational Testing Service.

Rock, D.A., and Werts, C.E. 1979. *Construct Validity of SAT across Populations—An Empirical Confirmatory Study.* College Entrance Examination Board Research and Development Report RDR 78-79, No. 5, Princeton, N.J.: Educational Testing Service.